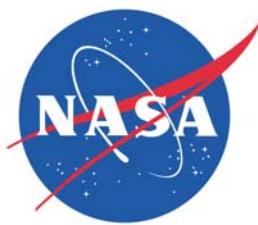


ICRAT 2010

Fourth International Conference on Research in Air Transportation

June 2010
Budapest, Hungary



**Fourth International Conference on
Research in
Air Transportation**



**1st – 4th June 2010
Budapest, Hungary**

Editors

Vu Duong
EUROCONTROL Experimental Centre
Centre de Bois des Bordes
BP 15
F-91222 Brétigny sur Orge CEDEX, France
Email : vu.duong@eurocontrol.int

David J. Lovell
1173 Glenn L. Martin Hall
College Park, MD, 20742, U.S.
Email: lovell@eng.umd.edu

ISBN 978-1-4507-1468-6

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks.

Copyright © 2010 ICRAT. All Rights Reserved

PREFACE

Welcome to the Fourth International Conference on Research in Air Transportation!

On the behalf of the ICRAT 2010 Steering Committee, we would like to express here our deep gratitude to the senior and young researchers in Air Transportation for having contributed to this young but challenging and exciting conference.

For this fourth edition of ICRAT, there were 94 qualified submissions by authors from 22 countries. The referee process resulted in 63 acceptances, for an acceptance rate of about 67%. All selected papers are of good quality, and we are very proud of the professionalism of all authors, reviewers, and of all Program Committee members. Thank you so much for your contributions and collaborations.

This is also the third time that Tutorials and a Doctoral Symposium have been included in the conference program. Three tutorials are scheduled for ICRAT 2010. There will be a full day tutorial on *Airborne Self Separation in Air Transportation*, with presentation by 10 of the leading researchers from Europe and the United States and half day tutorials on *Validation of ATM Operational Concepts* and on *Challenges Regarding the Integration of Unmanned Aircraft into Civil Airspace*. We are confident that they will increase the young scientists' understanding of "how things work" in air transportation. The Doctoral Symposium is expected to create a forum for young researchers to discuss their research approaches with senior researchers to obtain guidelines and support.

The opening session will have invited keynote speakers from the SESAR Joint Undertaking, from FAA, and from the Budapest University of Technology and Economics - all senior research scientists or strategists in Air Transportation. There will be two special keynote talks by senior air traffic controllers from Europe and the U.S. – both of whom have had extensive, close association with air transport research. We are very grateful for the presence, contributions, and support of these keynote speakers.

ICRAT 2010 and the proceedings you are handling are the result of much hard work from many people. We would like to thank:

- The authors and co-authors of the paper submissions. They are, of course, what makes the conference program great.
- The invisible tertiary reviewers, who often supply the most expert and informed comments on their review, and the ICRAT 2010 Program Committee. The 40 members on the committee spent most of their free time during the referee process to review the submitted papers and to return with careful comments. They are the guardians of the quality of the conference.
- The chairs of the committees: Dres Zellweger (General chair), Vu Duong and David Lovell (Program Chairs); John-Paul Clarke (Tutorial Chair); Mark Hansen (Doctoral Symposium Chair); and Sabrina Saunders-Hodge and Colin Meckiff (Grants & Awards Chairs).
- The logistics team led by Daniel Rohacs at Budapest University of Technology and Economics, and the conference secretariat team led by Yanjun Wang and Frizo Vormer of EUROCONTROL who worked hard to ensure the on-line processes with the authors, to collect, compile, and edit the final camera-ready proceedings.
- Telecom-ParisTech with the support to host the website, as well as for the time of Pr. Patrick Bellot and Loic Baud, who have worked pro-actively on the development and maintenance of the conference website.

- Daniel Rohacs and his local organizing committee members and volunteers, for all local arrangements, the printing of the proceedings on USB, and all the logistics at the conference place.
- The various institutions that provided the support for the paper process. The list includes the employers of all authors and co-authors and the employers of all reviewers and committee members.
- Eurocontrol, FAA, NASA, and JPDO for their financial support – which was instrumental in providing stipends for many of the ICRAT 2010 speakers.

Thank you all again, authors and reviewers, for your contribution to ICRAT 2010, that will surely be exciting. Thanks once more to the conference secretaries, Yanjun Wang and Frizo Vormer, and the principal local organizer Daniel Rohacs. The success of this conference will be yours!

Andres Zellweger and Jozsef Rohacs, *General Chairs*

Sabrina Saunders-Hodge, David Lovell, Vu Duong, Nicolas Durand, Colin Meckiff, *Steering Committee*

Chairs & Committees

			Program committee		
General Chair			Giovanni Andreatta	Uni. Padova	IT
Andres Zellweger	ATC Quaterly	US	Michael Ball	Uni. of Maryland	US
Jozsef Rohacs	BUTE	HU	Henk Blom	NLR	NL
Past Chairs			Marc Bourgois	EUROCONTROL	FR
George Donohue	Georges Mason Uni.	US	Steve Bradford	FAA	US
Antonin Kazda	Uni. Zilina	SK	Lorenzo Castelli	Uni. Trieste	IT
Vojin Totic	Uni. Belgrade	SB	Daniel Delahaye	ENAC	FR
Jean-Marc Garot	Ministry of Transport	FR	Dan DeLaurentis	Purdue Uni.	US
Program Chair			Nicolas Durand	DSNA	FR
Vu N. Duong	EUROCONTROL	FR	Hartmut Fricke	TU Dresden	DE
David Lovell	Uni. of Maryland	US	Norbert Fuerstenau	DLR	DE
Tutorial Chair			Mark Hansen	UC Berkeley	US
John-Paul Clark	Georgia Tech, US	US	Peter Hecker	TU Braunschweig	DE
Doctoral Symposium Chair			Jacco Hoekstra	TU Delft	NL
Marc Hansen	UC Berkeley	US	MingHua Hu	NUAA	CN
Grants & Award Chair			Antonin Kazda	Uni. Zilina	SK
Sabrina Saunders-Hodge	FAA	US	Diego Klabjan	Northwestern Uni.	US
Colin Meckiff	EUROCONTROL	FR	Paul Krois	FAA	US
Steering Committee			Bernard Lorenz	EUROCONTROL	HU
Andres Zellweger	ATC Quaterly	US	Peter Lindsay	Uni. of Queensland	AU
Sabrina Saunders-Hodge	FAA	US	Sandy Lozito	NASA Ames	US
Vu N. Duong	EUROCONTROL	FR	John Lygeros	ETH Zurich	CH
Nicolas Durand	DSNA	FR	Laura M. Major	Draper Lab.	US
Colin Meckiff	EUROCONTROL	FR	Colin Meckiff	EUROCONTROL	FR
Local Organizing Chair			Dritan Nace	UTC	FR
Daniel Rohacs	BUTE	HU	Sakae Nagaoka	ENRI	JP
Philippe Debels	EUROCONTROL	HU	Shinichi Nakasuka	Uni. of Tokyo	JP
Conference Secretary			Miquel Angel Piera	UA Barcelona	SP
Frizo Vormer	EUROCONTROL	NL	Wim Post	SESAR JU	BE
Yanjun Wang	Telecom-ParisTech	CN	Amy Pritchett	Georgia Tech	US
Webmaster			Jasenka Rakas	UC Berkeley	US
Loic Baud	Telecom-ParisTech	FR	Dirk Schaefer	EUROCONTROL	FR
			Jeff Schroeder	NASA Ames	US
			John Shortle	GMU	US
			Colin Smith	NATS	UK
			Senay Solak	UMASS	US
			Antonio Trani	Virginia Tech	US
			Seth Young	Ohio State Uni.	US
			Annalisa Weigel	MIT	US
			Yu Zhang	Uni. of South Florida	CN

Table of Contents

Preface	i
Chairs and Committees	iii
<i>Track 1: Advanced Modeling</i>	
Data Driven Modeling for the Simulation of Converging Runway Operations	3
A. Eckstein	
Tracking Failures in the Air Traffic System: An Ontology Based on Physical and Functional Decompositions	11
M. Gariel, E. Salaün, and E. Feron	
Game Equilibrium Analysis on an Auctioning Method for Single Airport Congesting Resource Allocation	19
F. Q. Liu and M. H. Hu	
Fair Slot Allocation of Airspace Resources Based on Dual Values for Slots	23
N. V. Pourtaklo and M. Ball	
Modeling and Predicting Taxi out Times	31
Y. Zhang, A. Chauhan, and X. Chen	
A TMA 4DT CD/CR Causal Model Based in Path Shortening/Path Stretching Techniques	37
C.A.Zúñiga, M.A.Piera, S. Ruiz, and I. Del Pozo	
Impact of Lightning Strikes on National Airspace System (NAS) Outages	45
A Statistical Approach	
A. Vidal and J. Rakas	
Ontology and Rules for International Airspace Security	53
M. R. Henriques	
Queueing Models for Operations in NextGen	61
T. Nikoleris and M. Hansen	
Applying Economy-wide Modeling to NextGen Benefits Analysis	67
K. Harback, L. Wojcik, Jr, M. B. Callaham, S. Martin, S. Tsao, and J. Drexler	
Optimal Route Generation with Geometric Recourse Model under Weather Uncertainty	75
Y. Yoon, M. Hansen, and M. Ball	
A Diffusion Approximation to a Single Airport Queue	85
D. Lovell, K. Vlachou, T. Rabbani, and A. Bayen	

Track 2: Airline Operations and Marketing

The Responses of Traditional Airlines to Low Cost Airlines	95
P. Srisaeng	
Estimating Domestic U.S. Airline Cost of Delay based on European Model	101
A. Q. Kara, J. Ferguson, K. Hoffman, and L. Sherry	

Track 3: Airport Design and Operations

Airport Ground Access and Egress Passenger Flow Model (AGAP)	111
M. Stefanik, B. Badanik, and M. Matas	
Operational Evaluation of an Airport Centered Flow Management	119
E. Rehwald, P. Hecker, and R. Kaufhold	
Door-to-Gate Air Passenger Flow Model	125
M. Matas and M. Stefanik	
The Airport Ground Movement Problem: Past and Current Research and Future Directions	131
J. A. D. Atkin, E. K. Burke, and S. Ravizza	
Modeling of Aircraft Surface Traffic Flow at Congested Airport Using Cellular Automata	139
R. Mori	
Iterative Planning of Airport Ground Movements	147
C. Lesire	
Potential of Dynamic Aircraft to Runway Allocation for Parallel Runways	155
M. Fritzsche, T. Günther, and H. Fricke	

Track 4: CNS/ATM

An Analysis of Delays in Air Transport in Japan	165
K. Kageyama and Y. Fukuda	
Analysis of “Tarmac Delays” at Philadelphia Airport	173
M. Medina and L. Sherry	
Trajectory Prediction by Functional Regression in Sobolev Space	179
K. Tastambekov, S. Puechmorel, D. Delahaye, and C. Rabut	
A New Method for Generating Optimal Conflict Free 4D Trajectory	185
N. Dougui, D. Delahaye, S. Puechmorel, and M. Mongeau	

Resource Allocation in Flow-Constrained Areas with Stochastic Termination Times and Deterministic Movement	193
M. Ganji, D. Lovell, and M. Ball	
The Air Traffic Flow Management Problem with Time Windows	201
L. Corolli, L. Castelli, and G. Lulli	
Flight Profile Variations due to the Spreading Practice of Cost Index Based Flight Planning	209
W. Rumler, T. Günther, H. Fricke, and U. Weißhaar	
Stochastic Integer Programming Models for Ground Delay Programs with Weather Uncertainty	217
C. N. Glover and M. Ball	
Enhanced Wind Magnitude and Bearing Prediction - Onboard algorithm	225
P. Krupanský, T. Neužil, E. Gelnarová, J. Svoboda, P. Mejzlík, and M. Herodes	
Radar Cross Section Generation of the Possible Non-cooperative Targets	229
R. Palme	
Study on Conflict Detection Method with Downlink Aircraft Parameters	237
A. Senoguchi and Y. Fukuda	
En Route Air Traffic Control Input Devices for the Next Generation	245
M. J. Mainini	

Track 5: Decision Support Tools

An Advanced Particle Filtering Algorithm for Improving Conflict Detection in Air Traffic Control	255
I. Lymeropoulos, G. Chaloulos, and J. Lygeros	
A Stochastic Model for Air Traffic Control Radio Channel Utilization	263
V. Popescu, H. Augris, and K. Feigh	
Impact of US Airline Network Topology on Air Transportation System Performance	271
T. Kotegawa, D. Fry, E. Puchaty, and D. DeLaurentis	
Estimation and Comparison of the Impact of Single Airport Delay to the National Airspace System using Multivariate Simultaneous Models	279
Y. Zhang, N. Nayak, and T. Diana	

Track 6: Environmental and Weather

Throughput/Complexity Tradeoffs for Routing Traffic in the Presence of Dynamic Weather	289
J. Krozel, J. S. B. Mitchell, A. Paakko, and V. Polishcuk	

Contribution of European Aviation on the Air Quality of the Mediterranean Region: A Modeling Study	297
J. Kushta, S. Solomos, and G. Kallos	
Generating Day-of-Operation Probabilistic Capacity Profiles from Weather Forecasts	305
G. S. Buxi and M. Hansen	
 Track 7: Future Concepts and Innovative Ideas	
Dynamic Allocation and Benefit Assessment of NextGen Flow Corridors	315
A. N. Zadeh, A. Yousefi, and A. Tafazzoli	
Coordinating Multiple Traffic Management Initiatives with Integer Optimization	323
A. M. Churchill, D. Lovell, and M. Ball	
Collaborative Rerouting in the Airspace Flow Program	331
A Framework for User-cost Based Performance Assessment	
A. M. Kim and M. Hansen	
An Optimisation Framework for Aircraft Operators Dealing with Capacity-Demand Imbalances in SESAR	339
L. Delgado and X. Prats	
High-fidelity Human-in-the-Loop Simulations as one Step towards Remote Control of Regional Airports - A Preliminary Study	347
C. Möhlenbrink, M. Friedrich, A. Papenfuß, M. Rudolph, M. Schmidt, F. Morlang, and N. Fürstenau	
Investigating String Stability of a Time-History Control Law for Interval Management	355
L. A. Weitz	
Towards Universal Beacon Code Assignment	363
Spatial and Temporal Analysis of En-route Traffic in NAS	
V. Kumar and L. Sherry	
 Track 8: Human Factors	
The Structure and Color Optimization Process to Generate Metro-like Maps of Flight Routes	371
C. Hurter, M. Serrurier, and R. Alonso	
A Participatory Design for the Visualization of Airspace Configuration Forecasts	379
N. Saporito, C. Hurter, D. Gianazza, and G. Beboux	
Flight experience and executive functions predict flight simulator performance in general aviation pilots	387
M. Causse, F. Dehais, and J. Pastor	

Economic Issues Provokes Hazardous Landing "Emotional" Neural Pathways	395
M. Causse, F. Dehais, J. Pastor, P. Péran, and U. Sabatini	
Empirical Analysis of Air Traffic Controller Dynamics	401
Y. J. Wang, F. Vormer, M. H. Hu, and V. Duong	
Predicting Controller Communication Time for Capacity Estimation	409
A. Vela, E. Salaün, P. Burgain, W. Singhose, J.-P. Clarke, and E. Feron	

Track 9: Prospective Studies and Economics

What Kind of Aviation Infrastructure Privatization is needed in China?	419
W. Lu and M. Finger	
Capturing the Impact of Fuel Price on Jet Aircraft Operating Costs with Engineering and Econometric Models	425
M. S. Ryerson and M. Hansen	
En route charges for ANSP revenue maximization	433
A. Violin, M. Labb��, and L. Castelli	

Track 10: Safety and Security

Bayesian Analysis of Accident Rate, Trend and Uncertainty in Commercial Aviation	443
E. A. Bloem and H. A.P. Blom	
Stochastically and Dynamically Coloured Petri Net Model of ACAS Operations	449
F. Netjasov, A. Vidosavljevic, V. Tasic, M. Everdij, and H. Blom	
A Quantitative Safety Assessment Tool Based on Aircraft Actual Navigation Performance	457
M. Vogel, C. Thiel, and H. Fricke	
Study of SESAR Implied Safety Validation Needs	465
J. J. Scholte, H. A.P. Blom, and A. Pasquini	
Collision Risk on Final Approach – A Radar Data Based Evaluation Method to Assess Safety ANP Based Obstacle Assessment Surfaces	473
C. Thiel and H. Fricke	
Comparison of Arrival Tracks at Different Airports	481
Y. M. Zhang, J. Shortle, and L. Sherry	
Stochastic Validation of ATM Procedures by Abstraction Algorithms	487
M. D. Di Benedetto, G. Di Matteo, and A. D'Innocenzo	

Doctoral Symposium

Establishing an Upper-Bound for the Benefits of NextGen Trajectory-Based Operations	495
G. Calderón-Meza and L. Sherry	
The Research of Multi-Airport Ground Holding Problem Based on the Schedule Optimization	501
B. Ye and M. H. Hu	
Using Online Data to Investigate Airline Passengers' Multi-Airport Choices	505
B. L. Luken and L. A. Garrow	
A Study of Characteristics of Solutions Obtained by Heuristics for Regional Air Traffic Flow Management	509
Y. Zhang and M. H. Hu	
A Mixed Integer Linear Model for Potential Conflict Minimization by Speed Modulations	513
D. Rey, S. Constans, R. Fondacci, and C. Rapine	
Using On-Line Data to Explore Competitive Airline Pricing Policies	517
S. Mumbower, and L. A. Garrow	
Evaluating Light Detection and Ranging (LIDAR) Technology in the Terminal Aerodrome Environment for Potential Enhancements and Air Traffic Management	521
A. E. Sinclair	
The Aircraft Sequencing Problem	525
N. Ahmad	
Models for Aircraft Landing Optimization	529
M. Mesgarpour, C. N. Potts, and J. A. Bennell	
Author Index	533

Track 1

Advanced Modeling

Data Driven Modeling for the Simulation of Converging Runway Operations

Adric Eckstein

The MITRE Corporation

Center for Advanced Aviation Systems Development

McLean, VA, USA

aeckstein@mitre.org

Abstract—A novel methodology is presented for generating data driven models for the general application of modeling and simulation. This approach relies on the use of principal component analysis to decompose a given data set into a basis of linearly uncorrelated modes. Data-driven models are then constructed from radar track data in order to develop models for a Monte Carlo simulation to evaluate the collision risk of converging runway operations.

Keywords - data driven models, converging runway operations, principal component analysis, modeling and simulation

I. INTRODUCTION

Conventional modeling approaches are frequently derived from performance specifications in order to infer conclusions from modeling and simulation. While this approach is often sufficient, it may rely heavily on assumptions or overly simplistic models that fail to capture operational variations in the data.

This study introduces a novel modeling approach in which data-driven models are generated to capture the operational variation directly from a given data set. This approach utilizes principal component analysis to reduce a data set to a series of linear models. This modeling technique is described in the context of a Monte Carlo simulation of converging and intersecting runway operations, used to generate collision risk factors and define the operational range for improved airport operations.

II. CONVERGING RUNWAY OPERATIONS

With increasing airport operations, new concepts are required in order to enhance throughput while maintaining a high level of safety. One such concept is reducing inter-operation time for dependent converging and intersecting runway operations. Specifically, this paper focuses on the arrival-departure converging runway operation. Fig. 1 shows an example of this operation at Chicago O'Hare International airport (ORD).

Conventional operations require "landing assured" for arrivals to 27R in order to release the departure from 32L, which limits the departure efficiency to that of arrivals on 27R. Alternatively, a "no-go box" defines a specified region prior to the runway threshold in which it is unsafe to release a departure. Provided there is no arriving aircraft in this region,

the departures would be cleared, increasing the departure efficiency. The goal is to define the smallest size and location of the "no-go box" to ensure safe operations while allowing the maximum throughput.

In this instance, we want to measure the risk associated with releasing a departure, given an arriving aircraft at a specified distance from threshold. The risk of the arriving aircraft initiating a missed approach and successively colliding with the departing aircraft should be sufficiently small to meet safety requirements. For this study, a collision was defined only as a function of the lateral trajectory, independent of altitude separation. Some of the factors that influence this risk are as follows:

- runway geometries and locations
- fleet mix
- missed approach rate
- missed approach initiation height

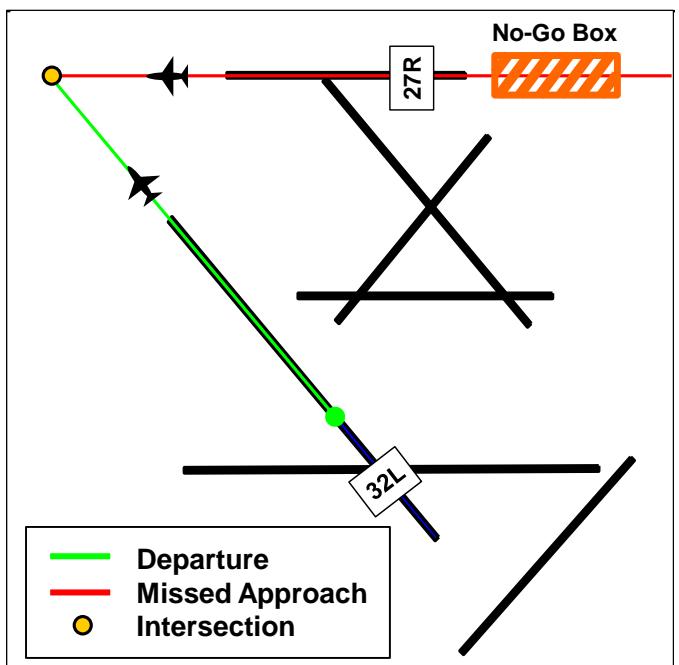


Figure 1. Converging runway operations at ORD.

- *departure clear-to-roll time*
- *aircraft trajectories*

A Monte Carlo approach has been adopted in order to compute the collision risk, where the airport operation is simulated using models which describe the variability in each component. Previous studies have adopted a similar methodology to assess the risk of these operations¹. Simulations of aircraft pairs are allowed to evolve independently, such that additional risk mitigations factors such as controller intervention or traffic collision avoidance system alerting are not requirements for the safety of the operation.

In this paper, the modeling approach for the aircraft trajectories is described. Since the aircraft trajectories define the separation time when the arrival/departure aircraft tracks cross, this is an essential component to modeling the risk. Each of the other factors is treated separately to determine the overall risk of a given operation, and is beyond the scope of this paper.

A. Radar Track Data

In order to capture the operational performance of aircraft, historical high-update radar track data is examined. The departure tracks were grouped by aircraft type for over a year of track data at 14 airports to generate enough statistics to capture the nominal behavior for a wide variety of aircraft types. Non-nominal events such as engine out, blunders, and navigational failure would require separate treatment and analysis. Aircraft types with insufficient data (less than 50 tracks) were alternatively modeled by a more generic grouping of tracks by engine type and approach speed category.

The high-update radar source data utilized reports aircraft position at 1 second intervals. The position and time of these track reports were passed through a series of least squares filters² from which the ground speed was estimated. Next, each departure was normalized to its throttle-up time, estimated for each speed profile. Each track was then reduced to the first 180 seconds of flight. The ground speed data set for each aircraft type then consists of 180 ground speed measurements for N departures. Fig. 2 shows a two-dimensional probability density function (pdf) plot of this data for the set of Airbus A319 aircraft. Each slice through the x-axis would represent a histogram of the track speeds at a given time. The mean as well as the 95% upper and lower confidence bounds on the speeds at each time is also shown.

III. DATA DRIVEN MODELS

A. Conventional Approaches

Trajectory modeling, in this context, relies upon defining a speed profile for a given aircraft type. While other factors such as climb rate and turn rate are integral components to a full trajectory model, the approach is initially defined for a single parameter. The methods to include such additional parameters is subsequently demonstrated. Using the runway location and heading, the speed profile can then be integrated in order to define a lateral trajectory for modeling and simulation.

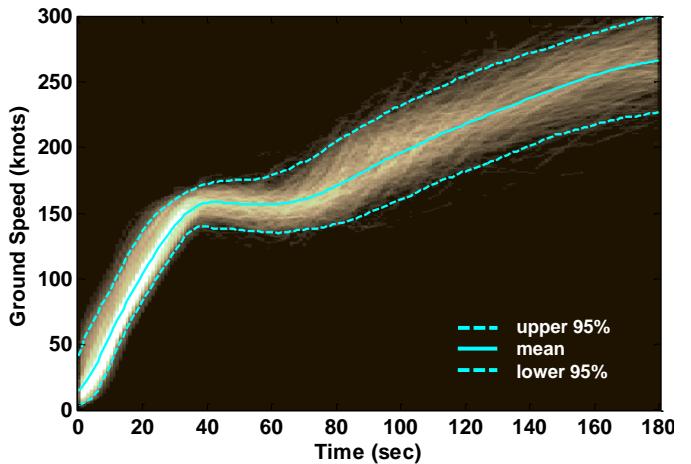


Figure 2. Two-dimensional pdf of A319 departure ground speed trajectory from track data .

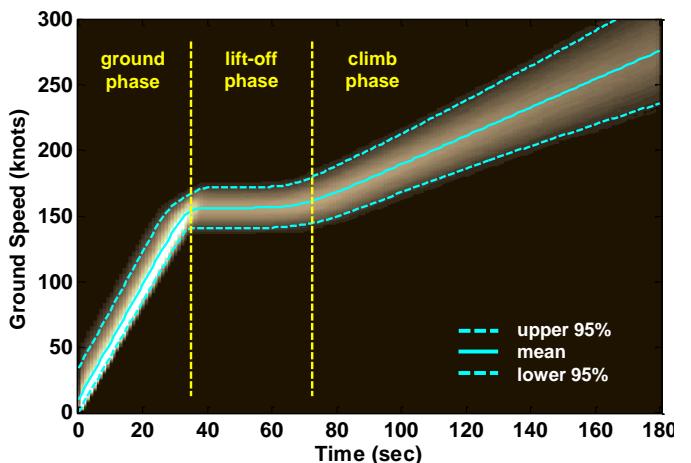


Figure 3. Two dimensinoal pdf of A319 constant acceleration trajectory model.

Numerous trajectory models exist for various aircraft types. These models often rely on aircraft and engine performance specifications, wind, and airspace models. However, such models often lack the operational variance in performance that would be seen in day-to-day operations. Variability may be estimated in the parameters which define such models but cannot guarantee that each model's variability is independent in defining the trajectory.

Alternatively, one can define the model directly from the given trajectory data. Conventional approaches to data-driven models might simplify the profile from Fig. 2 into simple linear segments with constant acceleration. Such an example is shown in Fig. 3. The speed profile consists of 3 segments. The first is a high acceleration ground roll phase, ending at the rotation point; followed by a constant speed take-off segment (during which gear and flaps are retracted); ending with a climb acceleration segment. Each segment is a simple linear model with coefficients drawn from known infinite distributions, based on fits to the given data.

While the bulk features of the speed profile are captured, there are apparent features in the actual speed profiles which indicate non-constant accelerations through these segments. In

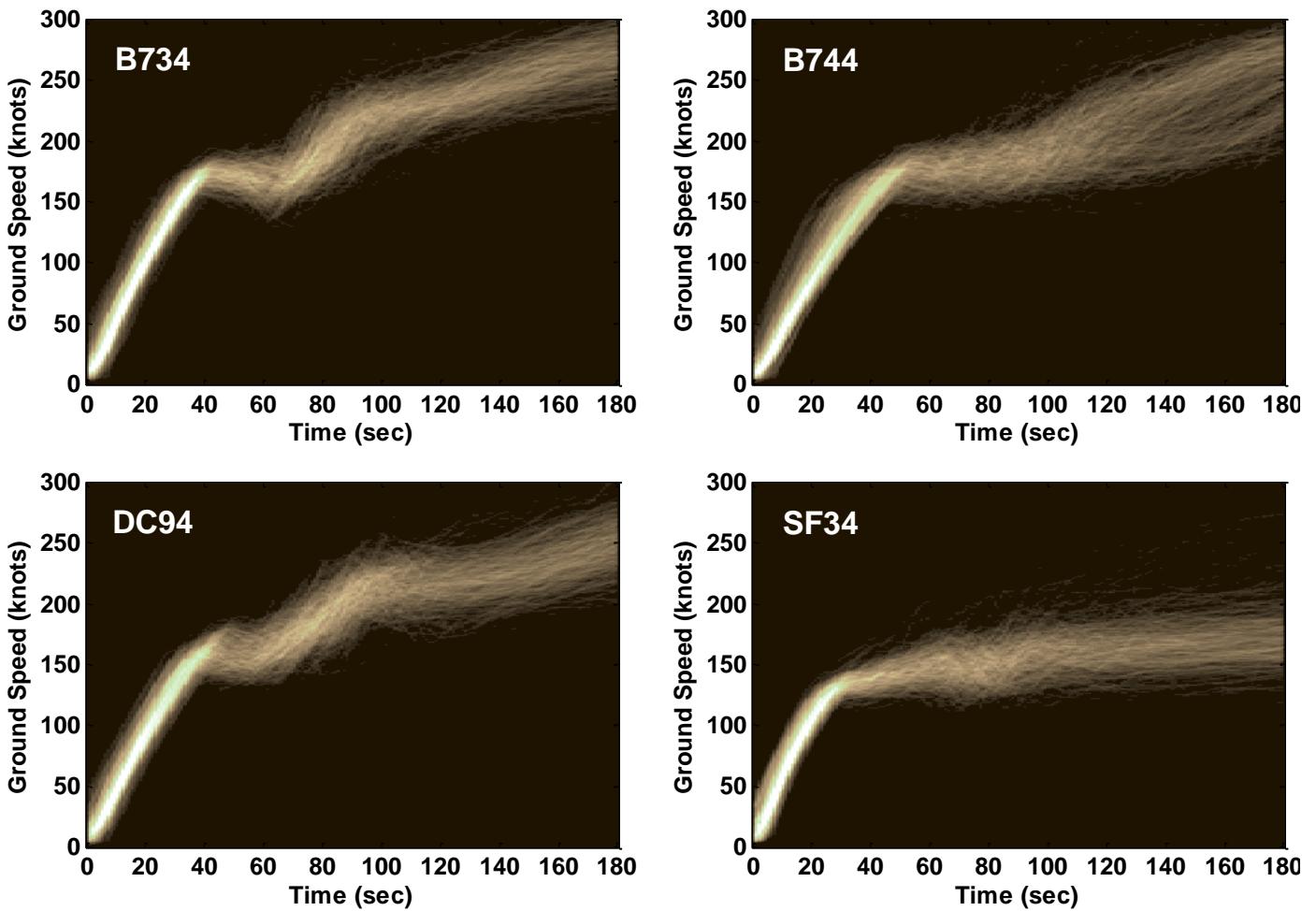


Figure 4. Two-dimensional pdf of departure ground speed trajectory for various aircraft types from radar track data.

addition, not all speed profiles demonstrated this basic 3-segment behavior, as shown by other aircraft types in Fig. 4. There are clearly different signatures in each speed profile as well as distinct differences in the variability, which cannot be captured from such a simple linear model.

B. Principal Component Analysis

An attractive alternative for data-driven models is the use of Principal Component Analysis (PCA). Previous studies have utilized PCA as part of data driven modeling to cluster data groups³ or deriving filter coefficients⁴.

PCA is a mathematical decomposition of a given data set through the use of eigenvalue decomposition⁵. Through PCA, a data set is decomposed into a set of uncorrelated principal components, which define a special basis set for the given data. PCA can also be described through the Proper Orthogonal Decomposition⁶ (POD) of a given data set:

$$\mathbf{X} = [\vec{x}_1 \ \vec{x}_2 \ \dots \ \vec{x}_n] = \mathbf{CM}^T = [\vec{c}_1 \ \vec{c}_2 \ \dots \ \vec{c}_k][\vec{m}_1 \ \vec{m}_2 \ \dots \ \vec{m}_k]^T \quad (1)$$

The data set, X , is defined by a series of random variables, x_i . In the context of this study, each random variable is the

ground speed at a given time (giving 180 random variables). The principal components (modes), m_i , are defined by the eigenvectors of the covariance matrix, $\mathbf{X}^T \mathbf{X}$. Similarly, the variance of each coefficient, c_i , is the eigenvalue of the covariance matrix. The distributions of these coefficients will be critical to the definition of the model.

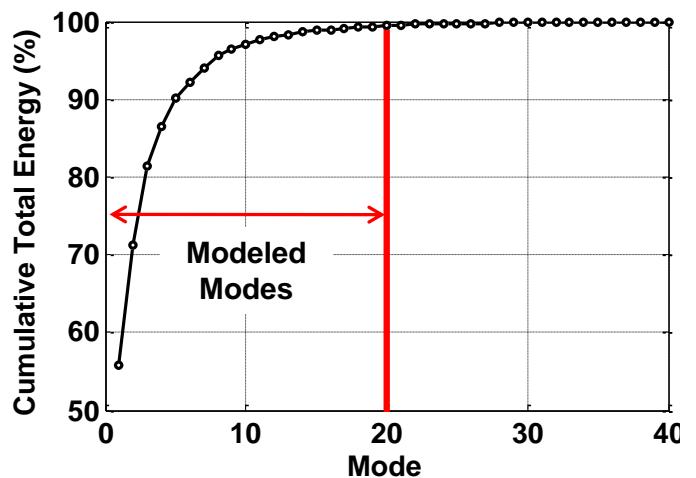


Figure 5. Cumulative total energy in the A319 PCA decomposition.

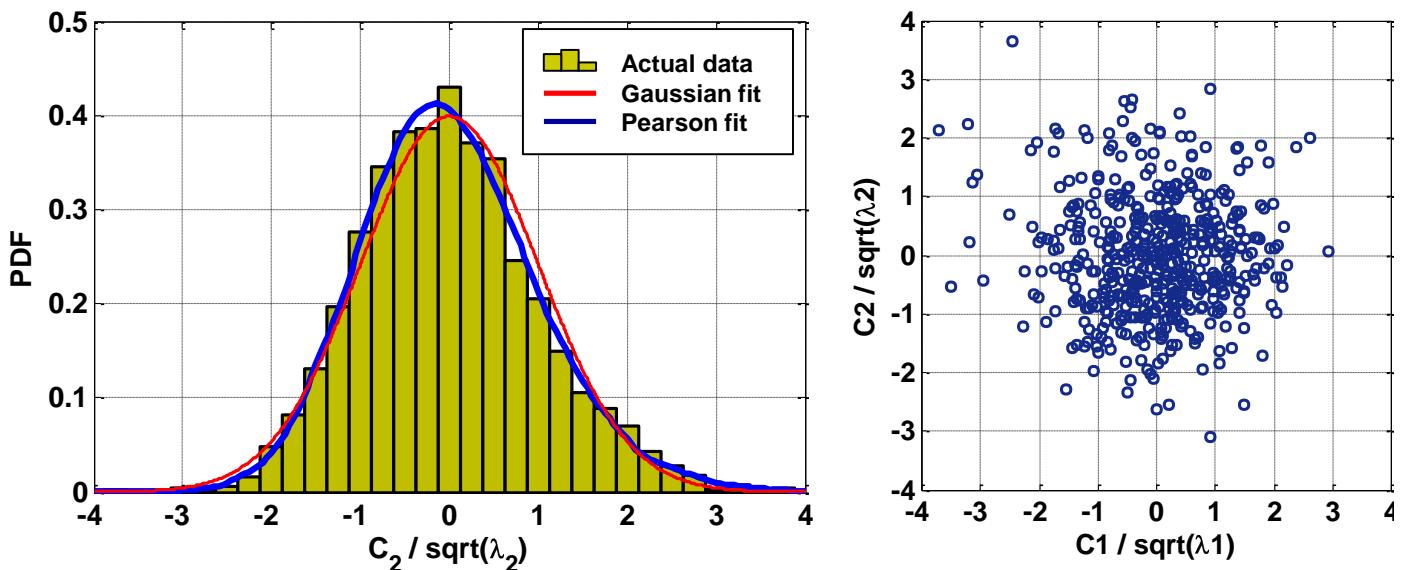


Figure 6. Continuous distribution fits to the second mode's coefficients (left) and scatter plot of first and second mode's coefficients (right).

The modes and coefficients represent a linear decomposition of the data set similar to a Fourier decomposition. However, each principal component in the PCA basis accounts for the maximum variability in the data. As such, the PCA basis will have less error than any other basis set when reconstructing using a subset of the components. Another essential property of the PCA basis is that each principal component is linearly uncorrelated with all other components.

C. Continuous Univariate Coefficient Distributions

For this example, the ground speed data set of the Airbus A320 aircraft is examined. PCA is used to decompose the set into its principal components (the mean mode is subtracted off prior to decomposition to center the coefficients). As PCA represents an optimal decomposition with respect to the variance of the data, it serves as a perfect tool to reduce the order of the data. Fig. 5 shows that by retaining the first 20 modes, the model is able to account for 99.5% of the variation in the data. While more or fewer modes could easily be retained, 20 modes are hereafter modeled for demonstration of the methodology.

Next, a model for the distribution of each mode's coefficients is developed. Fig. 6 (left) gives a histogram of the first mode's coefficient. Two continuous distributions are shown overlaid with the data, a Gaussian distribution and a Pearson Distribution⁷ (defined by mean, standard deviation, skew, and kurtosis). Relating this distribution back to the eigenvalue decomposition of the covariance matrix, the standard deviation of the Gaussian is equivalent to the square root of the eigenvalue. While the Gaussian is a good fit to the data, the Pearson fit better accounts for the skew in the data. For most well defined problems, the Pearson fit should be sufficient to capture the underlying distribution. However, certain problems may require alternative methods (such as Kernel smoothing), so the validity of the fit should always be verified with statistical tests.

This process is applied to each of the mode's coefficients, giving an infinite distribution for each mode. Dominant outliers were removed prior to fitting the infinite distribution in order to better approximate the data. Using the Kolmogorov-Smirnov test, each mode's coefficients was found to be indistinguishable from its Pearson fit at the 95% confidence level.

Fig. 6 (right) gives a plot of the second mode's coefficient against the first mode's coefficient (normalized to the square root of their eigenvalues). It is apparent by the random scatter of the data that the distributions are largely uncorrelated. This is again a result of the PCA, where each mode is linearly uncorrelated with other modes. This property enables sampling from each of the Pearson distributions independently, which can then be reconstructed into the speed trajectory by summing along each of the principal components. A sample trajectory is then given by:

$$\vec{v} = \sum_{i=1}^{20} s_i \cdot \vec{m}_i \quad (2)$$

where s_i is a random sample of the Pearson distribution from the i^{th} mode. The sample trajectory, v , is then used as an input to the Monte Carlo simulation. With any other basis decomposition, multivariate sampling would be required, which is much more complex and can lead to several additional sources of error.

In practice, bounds were placed on each coefficient and the variance of the coefficients in order to prevent unrealistic trajectories caused from sampling from infinite distributions. These bounds were determined from the sample size and the extrema observed in the actual data.

Fig. 7 shows the two-dimensional pdf of the PCA model for the Airbus A319 speed profile. Since the PCA basis is defined from the data set, the model is able to capture all of

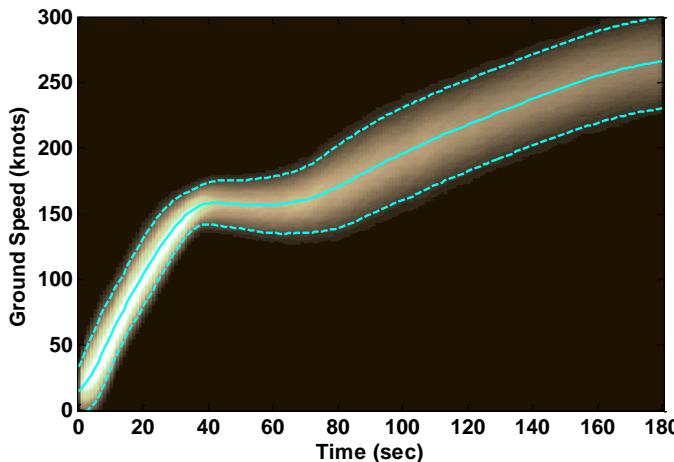


Figure 7. Two-dimensional pdf of A319 PCA decomposition trajectory model.

the effects of the input data set (shown in Fig. 2), without any *a priori* assumptions on the form of the model. This feature sets the PCA modeling approach apart from other conventional models. This model is also able to directly capture the correlation between speeds at different times, since each speed profile is a linear combination of the principal components of

the original data set.

Because this approach does not rely on any *a priori* inputs in order to construct the model, we can follow the same methodology for other aircraft types. The models for each of the aircraft types from Fig. 4 are shown in Fig. 8. Each of the models is able to capture the distinct signatures and variability from its corresponding data set.

D. Nonlinear Correlation Errors

One of the key assumptions in this modeling approach is the independence of each principal component from the others. In terms of the model, this is based on the assumption that the underlying variation in trajectories is caused by independent random variations. Each mode, by definition, is linearly uncorrelated with all other modes. However, this does define independence, as higher order correlations may exist in the data or multivariate groupings of tracks.

Fig. 9 plots a two-dimensional pdf of the ground speeds for 500 McDonnell Douglas DC10 aircraft. It is apparent that there is a largely bimodal variation in the data during the climb phase of the departure. This is likely caused by variations in airspeed restrictions, region, or aliasing of aircraft types. Regardless, we are attempting to define a generic DC10 trajectory model based upon the full data.

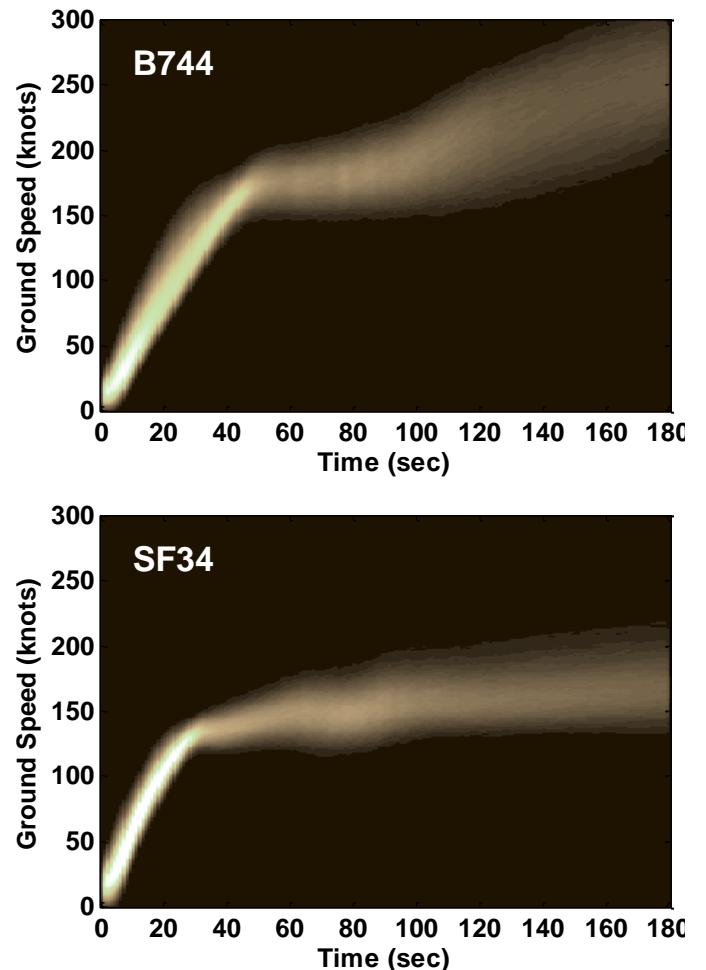
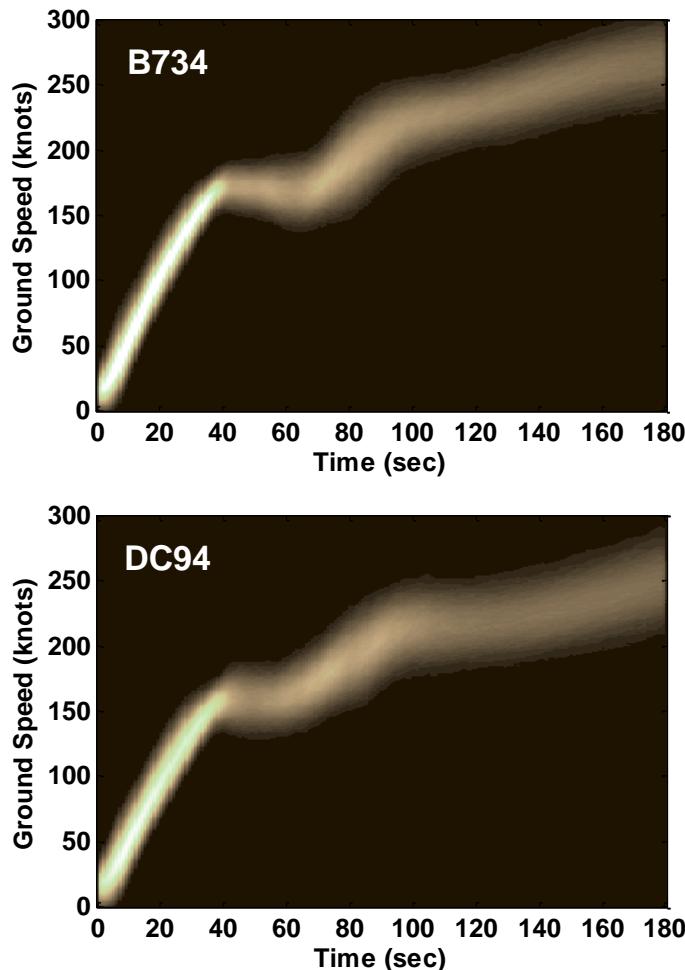


Figure 8. Two-dimensional pdf of PCA decomposition model for various aircraft types.

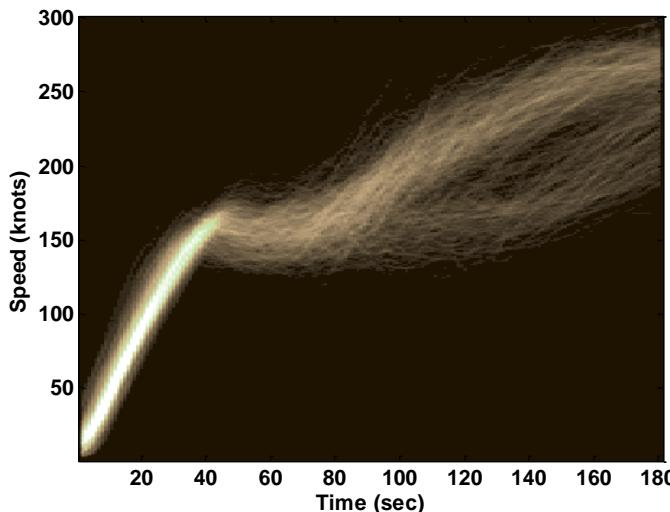


Figure 9. Two-dimensional pdf of DC10 departure ground speed trajectory from track data

Following the methodology described above, the PCA decomposition is applied to the DC10 tracks. Fig. 10 gives a plot of the first two coefficients, normalized to the square root of their respective eigenvalues (similar to the plot in Fig. 6, right). Although the two coefficients are linearly uncorrelated, there is clearly a higher-order correlation between the two. This correlation is directly related to the bimodal grouping apparent from Fig. 9. Fig. 11 (left) shows the end result when the model is applied to the bimodal distribution of tracks. The variability between the two groups is smeared out across the range of speeds, producing an inaccurate model. While a dominant source of this error is the approximation of the C1 coefficient by a unimodal distribution, even a perfect match to the bimodal pdf would still produce an erroneous model, since the principal components contain higher order correlations.

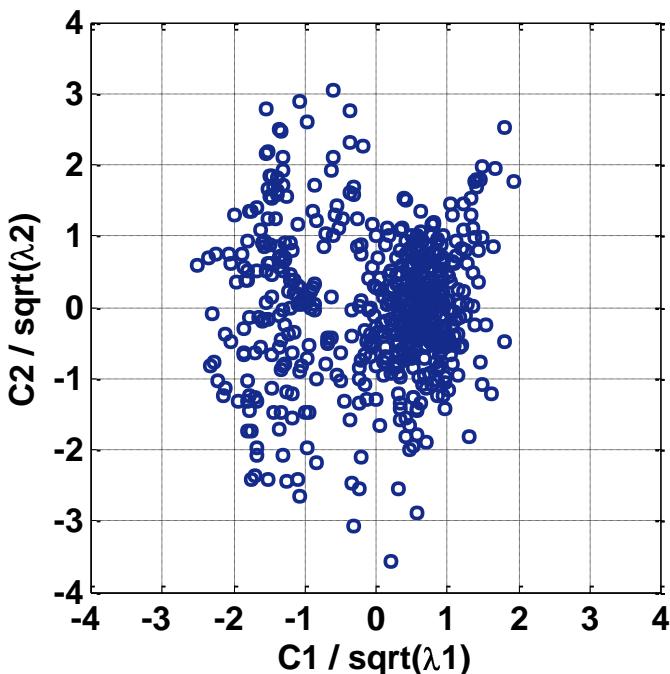


Figure 10. Scatter plot of DC10 first and second mode's coefficients.

Alternatively, we can start by dividing the set of DC10 trajectories by clustering the first couple modes' coefficients, from Fig. 10. Each of the groups is then independently modeled, creating two models for the DC10 aircraft. Random trajectories are then sampled from each of the two models, respective to their proportions in the data. This bimodal modal is shown in Fig. 11 (right), which is a much better model for the initial data set.

IV. VALIDATING THE TIME TO INTERSECTION

The time which it takes a departure to cross the missed approach trajectory is a key factor when applying the PCA model to the Monte Carlo simulations. The time it takes the departure to reach the intersection point is found by integrating the speed profiles until reaching a distance of 14,400 ft (the distance from the T10 taxiway of runway 32L to the intersection point). These times are determined for the recorded tracks and the PCA model, shown in Fig. 12 for the Airbus A319 aircraft. There is clearly a better match for the PCA model than the simpler constant acceleration approach.

V. EXTENDING THE MODEL TO ADDITIONAL DATA

Models are often required which can take a broader input in order to capture the relation between different parameters. In the context of the converging runway operations, the atmospheric conditions will have a substantial impact upon the performance of the aircraft, namely through the temperature and winds.

We can extend the PCA modeling concept to include these ancillary parameters as additional random variables in the PCA decomposition given in Eq. 1. The temperature and winds, measured on the ground at the airport during the release of the departure, are supplemented to each trajectory. Considering the wind a 2-component parameter, and the temperature a scalar, we now have an additional 3 random variables in the model. However, the data set, X , now consists of random variables with different units. Depending upon the scale of the data, the PCA will produce different results.

Each data set, X , can be normalized with respect to the variance of its input parameters. This is similar to the PCA obtained by standardized variables [5], where each variable is centered around its mean and normalized to the square root of its variance. In this context, we normalize groups of variables, such that the 180 ground speeds are normalized to their average variance, the 2 wind speeds are normalized with respect to their variance, and the temperature normalized to its variance:

$$\mathbf{X} = \begin{bmatrix} \frac{\vec{v}_1}{\sigma_v} & \frac{\vec{v}_2}{\sigma_v} & \dots & \frac{\vec{v}_{180}}{\sigma_v} & \frac{\vec{w}_1}{\sigma_w} & \frac{\vec{w}_2}{\sigma_w} & \frac{t_1}{\sigma_t} \end{bmatrix}$$

The PCA modeling approach can then be applied to \mathbf{X} where the last 3 elements of the resulting modes will contain the proportions of the principal components dedicated to the winds and temperatures. The modes are then parsed by their parameters and the normalization removed in order to return to the initial units of the data.

Utilizing this approach to capture the surface wind and temperature, arrival-departure pairs can be sampled and

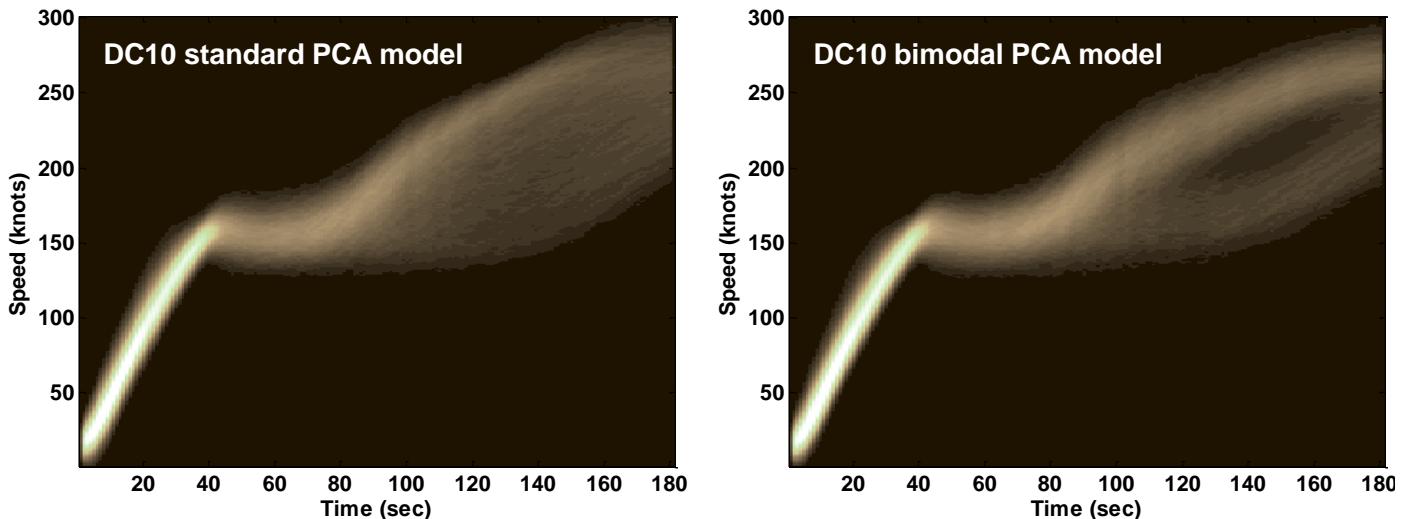


Figure 11. Standard PCA model of DC10 speeds (left) and bimodal PCA model of DC10 speeds (right).

matched according to the atmospheric conditions. Additional modes are included to account for the convergence of the energy of each parameter (similar to Fig. 5).

Conventional approaches might try to adjust the speed profiles or performance models based on the recorded surface measurements in order to account for these factors. However, such an approach adds a substantial number of assumptions which begin to degrade the accuracy of the model. By adding these components to the PCA model, no assumptions are made. Rather, aircraft pairs are selected which were shown to have similar surface conditions when they were flown.

VI. CONCLUSIONS

This paper presents a novel methodology for generating data driven models for the Monte Carlo simulation of airport operations. Specifically, this approach is applied to derive trajectory models for the application of converging runway operations collision risk analysis. In order to estimate the collision risk, it is necessary to have precise models of each

aircraft type's speed trajectory, which are obtained through data driven models from radar tracks.

Data sets were assembled using an aggregate of radar tracks from each aircraft type. Each data set was then decomposed using principal component analysis to obtain an empirical basis set of uncorrelated modes. The coefficients of each mode were then modeled using a Pearson distribution. Sample trajectories were then obtained by randomly sampling the Pearson distribution and summing over each of the modes in the PCA basis set. The resulting trajectories were shown to very closely match those of the input data set, subject to the assumption that the underlying data is governed by random independent fluctuations. Furthermore, it was shown that this approach can be extended to include additional parameters such as winds and temperature, which can be utilized to match the surface conditions of simulated aircraft pairs.

While this approach is provided in the context of the Monte Carlo simulation of converging runway operations, the methodology extends to a more general approach to modeling and simulation, capable of accurately capturing the relation across large data sets and between different data types, independent of the form of the input data.

ACKNOWLEDGMENT

The author would like to thank Joseph Clements for his reviews and contributions to the technical approach as well as the design of the Monte Carlo simulation methodology.

DISCLAIMER

This work was produced for the U.S. Government under Contract DTFA01-01-C-00001 and is subject to Federal Aviation Administration Acquisition Management System Clause 3.5-13, Rights In Data-General, Alt. III and Alt. IV (Oct. 1996).

The contents of this document reflect the views of the author and The MITRE Corporation and do not necessarily reflect the views of the FAA or the DOT. Neither the Federal Aviation Administration nor the Department of Transportation

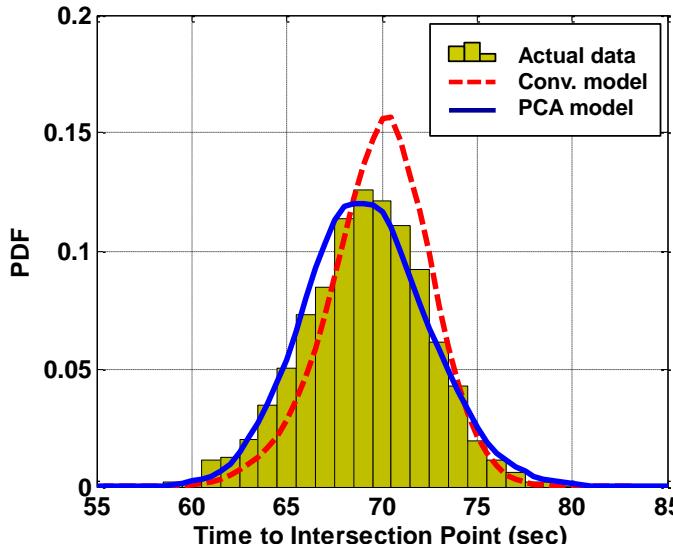


Figure 12. Departure time from T10-RW32L to the intersection point.

makes any warranty or guarantee, expressed or implied, concerning the content or accuracy of these views.

Approved for public release: 10-0130. Distribution Unlimited.

REFERENCES

- [1] H. A. P. Bloom, M. B. Klompstra, G. J. Bakker, "Accident Assessment of Simultaneous Converging Instrument Approaches," *Air Traffic Quarterly*, 2003, Vol 11(2), pp. 123-155.
- [2] A. Eckstein, "Automated Flight Track Taxonomy for Measuring Benefits from Performance Based Navigation," Proceedings of the 9th Integrated Communications Navigation and Survailence (ICNS) Conference, 2009, Arlington, VA.
- [3] W. Matsutik, H. Pfister, M. Brand, and L. Mcmillian, "A Data Driven Reflectance Model," *ACM Transactions on Graphics*, Vol 22, No 3, July 2003.
- [4] J. Hung, H. Wang, and L. Lee, " Comparative Analysis for Data-Driven Temporal Filters Obtained Via Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) In Speech Recognition," in Proceedings of ICASSP, 2002.
- [5] R. A. Johnson and D. W. Wichern, "Applied Multivariate Statistical Analysis 5th Ed.," Upper Saddle River, NJ, Prentice Hall, 2002.
- [6] Y. C. Liang, 2002, "Proper Orthogonal Decomposition and its Applications - Part I: Theory," *Journal of Sound and Vibration*, vol. 252, no. 3, pp.527-544.
- [7] K. Pearson, "Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material," *Philosophical Transactions of the Royal Society of London. A*, Vol. 186 (1895), pp. 343- 414.

Tracking Failures in the Air Traffic System: An Ontology Based on Physical and Functional Decompositions

Maxime Gariel, Erwan Salaün, and Eric Feron

School of Aerospace Engineering

Georgia Institute of Technology

Atlanta, Georgia 30332-0250, USA

maxime.gariel@gatech.edu, erwan.salaun@gatech.edu, feron@gatech.edu

Abstract—This paper presents an ontology for the air traffic system that aims at tracking failures and at measuring their impact on air traffic operations. This model is based on physical and functional decompositions of the air traffic system, which splits into facilities, aircraft, technologies, human operators, communication media, functions, tasks and operations. Possible failures are introduced at different levels of the decomposition and their consequences can be easily analyzed thanks to links between the blocks of the model. Two case studies illustrate how this model allows to anticipate the failures propagation and to find alternative solutions. A prototype implementation using MATLAB/Simulink is presented and illustrates the propagation of a secondary surveillance radar failure.

I. INTRODUCTION

The Air Traffic System (ATS) is a complex system that involves thousands of pieces of equipments, vehicles, facilities and people working together. The current system is aging and its modernization requires new technologies, automation and operations that should be compatible with the existing system. Weaknesses in the current centralized, voice-communication-based system include travel delays due to weather, safety and security breakdowns, the inability to adapt to new technologies such as uninhabited aerial vehicles, and a lack of dynamic adaptability in the face of disturbances and failures. The goals for the Next Generation Air Transportation System (NextGen) [1] in the United States and SESAR [2] in Europe focus on significantly increasing the safety, security, and capacity of air transportation operations, via new procedures and technological advances for all modes of air transport. In NextGen and SESAR, aircraft are expected to have a broader range of capabilities than today's and to support varying levels of total system performance via onboard capabilities and associated crew training. Many aircraft will have the ability to perform airborne self-separation, spacing, and merging tasks independently. Increased use of automation, reduced separation standards, Super-Density arrival/departure operations [3], and additional runways allow busy airports to move a large number of aircraft through the terminal airspace during peak traffic periods. However, in order to enable future capacity, NextGen will encompass novel technologies, vehicle types and operational concepts [3], and will ultimately bring

forth new types (or modes) of failures and disruptions. If unattended, these disruptions could result in severe setbacks for the NextGen and SESAR agendas and the health of the air transportation system as a whole. Tracking the propagation and the impact of failure gets always more difficult with the increasing complexity of the system.

The introduction of new technologies and new aircraft is not possible unless they have been certified with a very low failure tolerance, resulting in very few critical onboard failures. Nevertheless, some faults still occur but are often due to exogenous factors such as the bird strike that downed flight 1549 [4] in the Hudson river. Ground infrastructures are also regularly affected by unexpected exogenous factors. On April 23, 2009, the Atlanta Hartsfield Jackson International Airport control tower was hit by a lightning and severe storms knocked out power to the area and the airport lights [5]. The tower had to be evacuated, leaving the airport inoperative, no aircraft being able to take off nor land. Probably more critical was the evacuation of the Southern California TRACON (SCT) in 2003 because of wildfires [6] threatening the facility. Technologies, or pieces of equipment are also subject to failures such as radar outages: in May 2007, the SCT was affected by an outage that let the controllers mapless for an hour [7]. Similarly, in 2004, a computer glitch in the radar system disabled the surveillance of flights above 24,000ft [8] in the United Kingdom. Flight data had to be entered manually, resulting in a decrease in capacity and an increase in spacing distances. Air Traffic Controllers (ATCs) are the eyes of the pilots and require a good sight to maintain the safety of the airspace. Surveillance is critical but cannot be achieved without reliable communications, which are used to transmit surveillance data from the radar to the control facilities, between ground facilities and aircraft, and among ground facilities. If a control center loses its communication capacities, tens or hundreds of aircraft are left deaf and blind. On September 25, 2007, Memphis Air-Route Traffic Control Center (ARTCC) experienced a total communication failure. Controllers had to coordinate with other ARTCCs using their cellphones [9], [10], [11], [12]. The breakdown lasted for about 4 hours. Communications are also critical for navigation purposes. If the frequencies carrying

the GPS signals are jammed [13] or spoofed, GPS navigation is not possible anymore. The lack of responsiveness of flight 188 [14] is an interesting example of communication and navigation failures due to a human error, when the pilots did not contact ATCs for over an hour and a half and overshot their destination airport by 90 NM.

In all of the examples above, disastrous consequences were avoided thanks to the extraordinary ability of humans to accommodate to unexpected situations. With the increase of automation in the Air Traffic Management (ATM) system, it becomes more difficult to track the impact of a partial or total failure. Sheridan analyzed the issues [15] associated with the human-automation interactions in the next generation air transportation systems: *“Because of the greater interconnectedness of aircraft and subsystems, equipment failures and misapplied procedures can cause perturbations that cascade throughout the whole system.”* [15] In the event of a failure of the automation, degradation mode should be available for the human controller to safely handle the system. A thorough knowledge and modeling of the degradation modes of the ATS is necessary to ensure its safety.

This paper presents an ontology of the ATS based on a multidimensional decomposition, aiming at analyzing the impact of failures. A large share of the work in ATM is devoted to improving the performances of the current system and assessing new concepts of operations. Being at the center of air traffic operations, ATCs have been often modeled ([16], [17], [18]). Human in the loop simulations are used to validate new concepts [19] and tools to identify human errors in air traffic control have been developed [20]. At a higher level, Pinon et al. modeled the air transportation system as a supply chain [21] to measure its performances and constraints. Using a modeled network of airports they studied the benefits of 4-D trajectory-based operations. Pinon et al. also presented a morphological decomposition of the air traffic operations, to evaluate and select airport technologies [22]. The system was decomposed from traffic phase, to possible improvements, to operational concepts, to functions, to technologies and finally, to sub technologies. A matrix of alternatives is created from this morphological analysis [23]. This decomposition is oriented towards finding appropriate technologies to optimize operations. However, this decomposition is unidirectional and does not allow to keep track of failures and to measure the loss of performances. Di Benedetto et al. modeled the ATMS as a stochastic hybrid system [24] to detect faults and mainly non-deterministic human errors due to lack of “situational awareness” [25]. Those hybrid systems allow to account for the dynamics of the agents and their potential errors. This ontology proposed in this paper does not include system’s dynamics but rather focuses on the deterministic health of the overall system.

Processes to study “Complex, Large-Scale, Interconnected, Open, Sociotechnical” (CLIOS) systems have been developed by Sussman [26]. The ATMS possesses the internal complexity [27], that is, the number of components in the system and their interactions, as well as the behavioral complexity [27],

that is the type of behavior that emerges from the system due to the components interactions, of CLIOS systems. The emergent behavior is difficult to predict but unlike in CLIOS systems, despite their complexity, the relationships between the subsystems of the ATMS are known and can be modeled. The use of CLIOS system analysis methodologies focuses on identifying policies or management intervention to improve the systems. If the ATMS was to be studied in the framework of a CLIOS system, this ontology would be a key enabler towards understanding its complexity.

In a common report [28], the FAA and Eurocontrol presented the safety techniques used in ATM. Techniques used to model the impact of failures on operations include, but are not limited to bow-tie analysis, even trees and fault trees [29], [30]. In all those frameworks, a failure or an error is first identified, and then, the possible causes (i.e the branches of the tree) need to be generated by an expert. This ontology is an automated way to generate fault trees.

The objective of this paper is to present a model that captures how failures or perturbations cascade throughout the system and that measures their impact in terms of loss of capabilities. If the impact of a failure is known, it becomes easier to ensure the graceful degradation of operations when the failure will occur. A “graceful degradation” of air traffic operations is defined as the smooth transition from nominal to degraded modes of operations [31]. For instance, previous works [32], [33] have presented algorithms and maps that enable a graceful degradation by spreading out the traffic in the event of a degradation in the ATS.

The remainder of this paper is organized as follows: section II introduces the ontology, its objectives, how the ATS is decomposed, the links between the blocks, and how failures are introduced and tracked. Section III consists of case studies to illustrate some examples of use for the ontology. Section IV presents a prototype implementation of the ontology before the concluding remarks.

II. AN ONTOLOGY FOR FAILURE TRACKING IN AIR TRAFFIC SYSTEMS

This section presents an ontology that models the ATS and allows to track failures in it. The objectives of the ontology, the modeled system, a description of the ontology, the influence structure and finally the failure propagation mechanism are presented.

A. Presentation of the Ontology

1) Objective of the Ontology: The objective of the ontology is to provide a better understanding of the propagation of failures in the ATM system, and to measure their impact in terms of loss of capabilities. The ontology shows the propagation of failures, from a facility, a controller or a technology, all the way to operational capabilities. The ontology allows the identification of alternate or backup technologies, to analyze how a loss of automation can be handled by a human controller to ensure the safe transition from a nominal

and automated, mode of operation to a degraded and manual mode of operation.

2) *System*: The modeled system is the air traffic system, consisting of all the infrastructures, technologies, communication media, people, etc, that are necessary for the air traffic system to be fully operative. The system also includes aircraft and pilots. The introduction of new technologies and automation systems can be tested and added to the ontology as they are being developed.

3) *Ontology description*: The ontology combines a physical decomposition of the major components of the ATS, and a functional decomposition of the air traffic operations into tasks and then functions. This ontology combines elements from the decompositions presented by Pinon et al. on the one hand ([22]), and Kim et al. on the other hand ([34]). Pinon et al. decomposed air traffic operations to identify enabling technologies. Kim et al. proposed a task decomposition for function allocation.

This ontology starts from a physical decomposition of the system in facilities and aircraft, then decomposes them into technologies and human operators. Human operators and technologies execute functions that are enabled by other functions and communication media. Then, those functions are used to execute tasks. Finally, tasks are combined together to enable operations. This later part, is the functional decomposition: Operations are decomposed into tasks, that are themselves decomposed into functions.

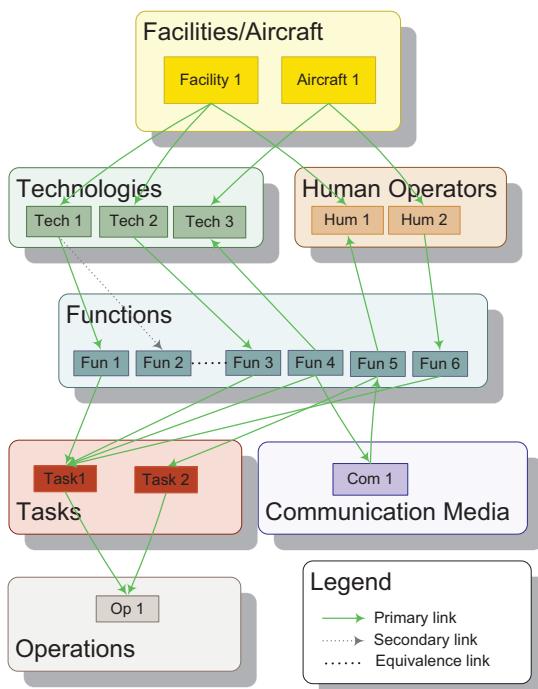


Fig. 1. Ontology's decomposition

Figure 1 presents a diagram of the decomposition of the elements of the ontology. The terms used in this model are defined as follow:

- **Facilities/Aircraft:** This category groups physical pieces of equipment and/or people located at the same place. A *Facility* refers to a building or place that provides a particular service or is used for a particular purpose. For instance, the TRACON facility refers to the physical building in which air traffic controllers work to direct aircraft in the corresponding TRACON airspace. A facility can also refer to a simple building, e.g. the building and mount for a radar or an Automatic Dependent Surveillance - Broadcast (ADS-B) ground station.
- **Aircraft:** An *Aircraft* refers to a vehicle that can fly and enter the controlled airspace, such as an airplane, a helicopter or an unmanned aerial vehicle.
- **Technologies:** A *Technology* refers to a physical piece of equipment such as a transponder, a radar, etc. A technology is located in a facility or an aircraft and executes one or several functions.
- **Human Operators:** A *Human Operator* refers to a human being qualified to execute the tasks required by his position/job. Human operators include pilots, air traffic controllers, dispatchers, etc. Human operators are located in ground facilities or an aircraft and execute one or several functions.
- **Communication Media:** A *Communication Medium* refers to the transmission channel or tool used to deliver information, such as air waves in a given range of frequencies, phone lines, etc.
- **Functions:** A *Function* refers to “a capability without a goal”, of a technology or a human being. Transmitting information or displaying information on a screen are examples of functions.
- **Tasks:** A *Task* refers to a tangible activity with a goal. A task is made possible through the combination of functions. Monitoring aircraft position for safe separation is an example of task.
- **Operations:** An *Operation* refers to a tangible activity with a goal resulting from the combination of several tasks. For instance, sequencing and merging is an operation that requires air traffic controllers to direct aircraft, pilots to follow ATC instructions and fly the aircraft.

This decomposition enables the introduction of failures at different levels (Section II-B). The propagation of a failure can be tracked in the ontology using its influence structure.

4) *Influence structure and dimension*: The influence structure of the ontology is the set of relationships and links that exists between the different components of the model. The signification of the links between the elements is presented in Table I. The dimension corresponds to the type of relationship existing between the linked blocks. The term *origin* refers to the block at the tail of the arrow, and *destination* refers to the block located at the head of the arrow. The relationship “*Hosts*” means that the destination block is located inside the origin block. The relationship “*Executes*” means that the origin block executes the destination block. The relationship “*Emits*” means that the origin block emits information using the destination block. The relationship “*Transmits*” means

TABLE I
INFLUENCE STRUCTURE AND DIMENSIONNALITY

Origin	Destination	Dimension	Meaning
Facility	→ Technology	Hosts	The technology is physically located inside the facility.
Facility	→ Human operator	Hosts	The human operator is physically located inside the facility.
Technology	→ Function	Executes	The technology executes this function. The information available to the technology is used to perform the function that will generate new information.
Function	→ Technology	Provides information	The output of this function is used by the function. The information generated by the function is used by the technology.
Human operator	→ Function	Executes	The human operator executes this function, generating new information.
Function	→ Human operator	Provides information	The human operator uses the output of this function. The information is received by the operator.
Function	→ Communication medium	Emits on	The output of the function is transmitted over the communication medium. The communication medium must be available for the information to be successfully transmitted.
Communication medium	→ Function	Transmits	The communication medium transmits information that can be captured by the receiving function.
Function	... Function	Is equivalent	The two functions are equivalent, in terms of role. They might have a different level of performance.
Function	→ Task	Enables	The function enables the task. A task might require several functions to be achieved.
Task	→ Operation	Enables	The accomplishment of the task is required for the operation to be conducted.

that the origin block transfers the information to the destination block. The relationship “Enables” means that the origin block makes the achievement of the destination block feasible. The relationship “Equivalence” does not carry any dependence information. It is used to determine redundancy in the technologies. The dimension is inferred by the nature of the elements on both sides of the relationship.

The ontology has three types of links: primary, secondary and equivalence.

- **Primary links:** Primary links correspond to nominal interactions between the different components. They are represented by colored arrows: a green and plain arrow indicates a link working nominally. A dashed orange link indicates that some of the information nominally carried by the link is missing. A dotted red arrow indicates the link is no functional.
- **Secondary links:** Secondary links correspond to redundancies, not used in nominal modes. They are also represented by colored arrow: when the link is inactive, it is represented by a dashed gray line and when active, it takes the colors of a primary link. For instance, the primary radar can be used as a backup for the secondary radar, but does not provide the same level of performance. The functions enabled by the primary radar are contained in the model, but the links are listed as secondary, since they are not used during nominal operations.
- **Equivalence links:** Equivalence links join blocks with similar characteristics. They are represented by black dotted lines. Two technologies are equivalent if and only if they are identical. If they are not, they can perform identical functions which will have the equivalence relationship. Two equivalent functions can have different

level of performance, which are indicated in the description of the function. Equivalence links allow the ontology to find redundant systems to perform failed functions.

B. Failures and degradations modes

The ontology enables the introduction of failures at all the levels of the decomposition. Failures can affect single or multiple blocks but cannot be introduced on links. The color of the links presented in the previous section refers to the type and the availability of the information they carry. The coloring is only a consequence of a failure being propagated. Failures can affect:

- **Facilities/Aircraft:** Failures affecting facilities and aircraft are potentially the most difficult to handle, since they host many people and technologies. Such failures can be total or partial. When a facility failure is total, it is propagated to technologies and human operators located in this facility. When the failure is partial, only some technologies or human operators will be set as “failed”. A total failure can be visualized as a master switch for all the technologies and people in the facility. A partial failure can be seen as a switch for a particular room. This is captured in the model by switching to inoperative only some elements (technologies or human operators) located inside this facility.
- **Technologies:** A technology can fail because the facility in which it is located fails, or because the technology itself fails. The same way facilities fail, technology failures can be partial or total. If the failure is total, all the functions enabled by the technology will be set to inoperative. If the failure is partial, only a set of functions will be set to inoperative.

TABLE II
EXAMPLE 1: PROPAGATION OF A FAILURE DUE TO INOPERATIVE BAROMETRIC ALTIMETER

Origin (Type)	Destination (Type)	Explanation	Level of Failure
Barometric altimeter (Tec)	→ Measure Altitude (Fun)	The altimeter cannot measure the altitude.	Total
Measure Altitude (Fun)	→ Onboard Mode-S Transponder (Tec)	The Mode-S transponder cannot get the altitude information.	Total
Onboard Mode-S Transponder (Tec)	→ Transmit Information (1090MHz) (Fun)	The Mode-S transponder cannot transmit the altitude information.	Partial
Transmit Information (1090MHz) (Fun)	→ Radio Waves (1000MHz) (Com)	There is no altitude information to transmit.	Partial
Radio Waves (1000Mhz) (Com)	→ Receive Information (1090MHz) (Fun)	There is no altitude information to receive.	Partial
Receive Information (1090MHz) (Fun)	→ Ground Mode-S Transponder (Tec)	The transponder cannot receive the altitude information.	Partial
Ground Mode-S Transponder (Tec)	→ Display of aircraft position (Fun)	The position of the aircraft cannot be accurately displayed since the Mode-S transponder did not receive altitude information.	Partial
Display of aircraft position (Fun)	→ Surveillance (Task)	The surveillance task cannot be executed properly as the altitude of an aircraft is missing.	Partial
Measure Altitude (Fun)	→ Fly holding Pattern (Task)	It is not possible to fly a holding pattern since it requires to maintain the altitude, which is not available.	Total
Fly holding Pattern (Task)	→ Sequencing and Merging (Ope)	Sequencing and merging might require an aircraft to fly a holding pattern. Since this task cannot be achieved by all aircraft, this operation runs in a degraded mode.	Partial
Surveillance (Task)	→ Sequencing and Merging (Ope)	Sequencing and merging require that the surveillance task is achieved properly.	Partial

- Human Operators:** Failures affecting human operators are modeled the same way as failures affecting technologies.
- Communication Media:** When a communication medium fails, the information it carries cannot reach its destination. Therefore, the link exiting the medium will be disabled, meaning that the information cannot be transmitted.
- Functions:** A failure cannot be introduced at the function level. If a technology or an operator cannot execute a function, it is modeled as a partial failure of the technology or operator. A function can fail if its input link(s) carry failures.
- Tasks:** Tasks can fail by the propagation of functions failures. Failures at a task level can also be introduced to model human errors. Task failures propagate to the operation level. Backtracking of task error is possible but likely to provide too many possible origins.
- Operations:** Operations can fail by the propagation of tasks failures.

When a failure is introduced in the ontology, the failure is propagated along and its impact can be measured by an incapacity of executing tasks and operations. Since links also carry partial failures, the model also allows its user to measure decrease in performances.

III. CASE STUDY

This section presents two case studies to illustrate some uses for the ontology. The first case illustrates the propagation of a failure of the barometric altimetry [35].

The second case shows how the ontology can be used to find alternative technologies in the event of a GPS jamming.

A. Failure of the barometric altimetry in one aircraft

In this example, a failure is introduced in the barometric altimetry of an aircraft. Such an example has been previously studied [35] for evaluating future concepts of operations. Agent based simulation were run to analyze the impact on operations and the loss of capabilities, from an aircraft-based point of view. It is assumed that the technologies providing this function are inoperative. Figure 2 presents a simplified version of the model. The blocs are depicted in red if they are the origin of the failure, or if this element has failed totally. A block in orange is partially affected by the failure. It is visually easy to follow all the elements that completely failed and those which suffer of a loss in capabilities. Table II explains how the failure propagates along the ontology. The diagram in Figure 2 is a simplified version of the trajectory of the failure through the model, as it does not present all the elements of the system. Only some blocks were selected to illustrate the example.

B. Jamming of the GPS signal

Future operations highly rely on accurate positioning using GPS. Super-Density Operations [3] will require aircraft to precisely follow predetermined trajectories consisting of a sequence of way-points coordinates. GPS is necessary to ensure Required Navigation Performance (RNP) operations. ADS-B literally depends on the aircraft being able to determine its position, in order to broadcast it to ground stations and to other aircraft. Figure 3 depicts the impact of a GPS jam [13]

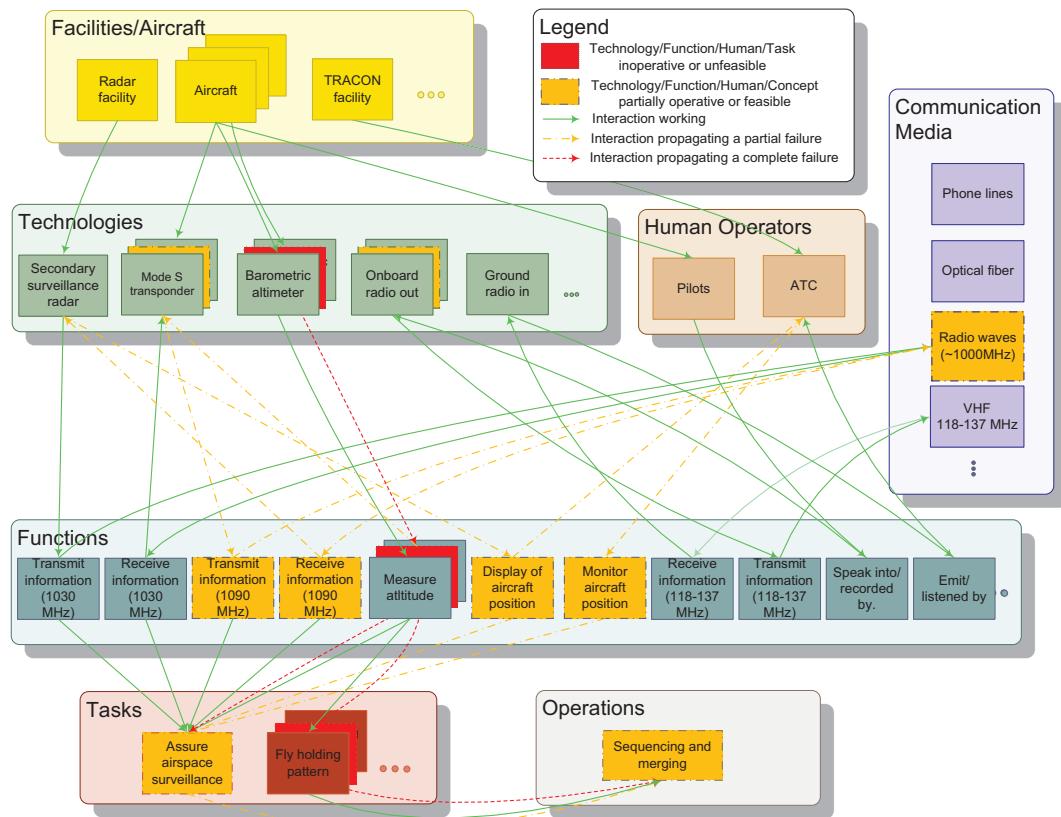


Fig. 2. Propagation of an aircraft's barometric altimeter failure through this ATS

on ADS-B operations. This representation is slightly simplified and is organized by entity for a more compact view. This example presents only one direction of communication, that is only aircraft 2 trying to determine the position of aircraft 1 using ADS-B. This example shows how Traffic Information Services (TIS) could be used as a backup to ADS-B in operations in terminal areas. The link between ADS-B out of aircraft 1 and ADS-B in of aircraft 2 is the primary link for aircraft 2 to obtain surrounding traffic's information. If this link fails, the secondary link is activated and TIS is used as a backup system.

IV. IMPLEMENTATION

This section presents a implementation of the ontology using MATLAB and Simulink. To show the capabilities of the ontology, a limited ATMS was simulated. Matlab and Simulink were chosen for the rapidity of prototyping. Simulink offers a great simulation environment with interaction and communication between elements. The ontology was implemented in Simulink, and a control panel module to introduce the failures and track their impact was created using MATLAB.

A. Simulink Model

In this implementation, yellow blocks represent facilities and aircraft. When going inside those boxes, appear the technologies and human operators hosted by those facilities.

The outputs of the technologies and operators are the functions accomplished.

B. Matlab Interface

An automated MATLAB interface was developed together with the Simulink model. This interface automatically reads the Simulink model and looks for all the facilities, etc, and the functions enabled. Figure 5 presents the control panel that enables the introduction of failures. By checking one or several boxes, the corresponding components will be set as inoperative. Figure 6 presents the control panel that shows the status of communication medias, tasks and operations. In both control panels, a green background indicate a nominal state. A red background indicates that the function/task/operation can not be executed.

Figures 5 and 6 present a scenario where the secondary radar is inoperative. The cascade of failures affects principally the controller that cannot monitoring traffic since its control station does not display aircraft's positions. Aircraft's mode-C transponder will not broadcast the aircraft's altitude since there were no such request, the secondary surveillance radar (SSR) being inoperative. The impact on tasks and operations is major since all activities requiring air traffic control feedback cannot be executed anymore. The aircraft are left on their own.

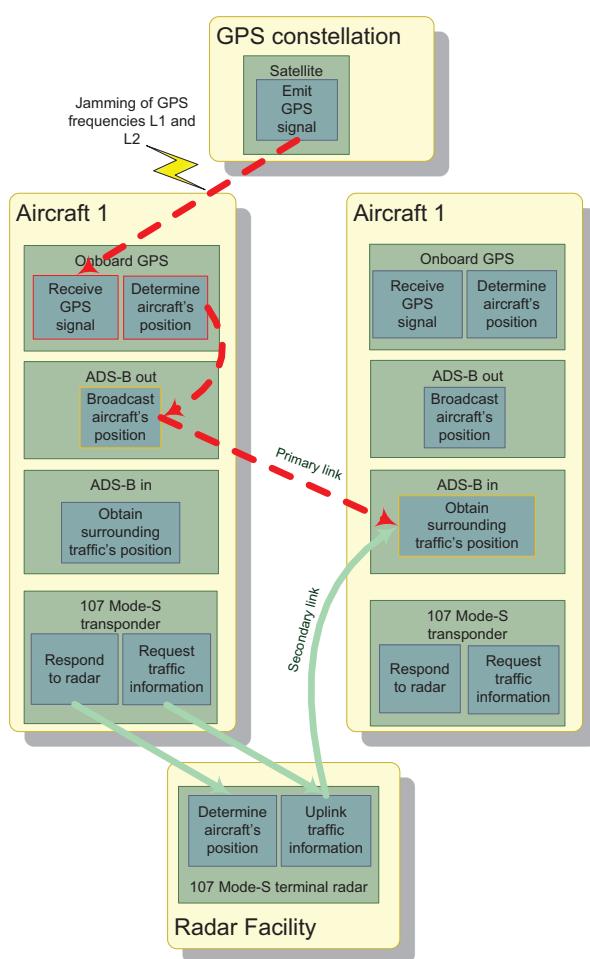


Fig. 3. Impact of GPS jamming on ADS-B operations

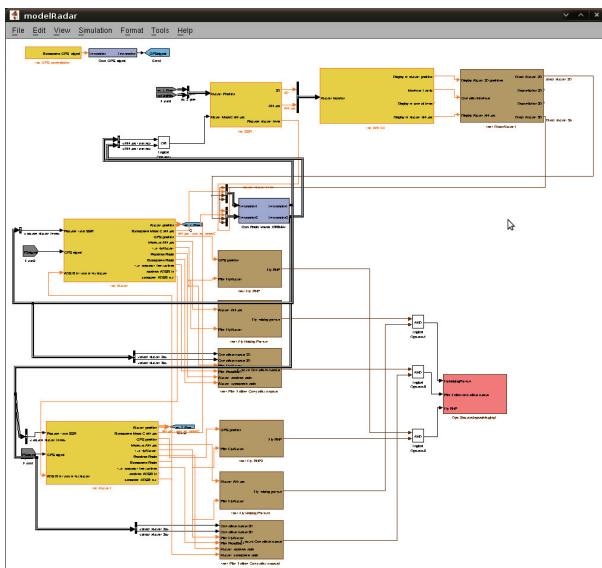


Fig. 4. Simulink model for the ATS Ontology



Fig. 5. MATLAB interface - Technologies control panel for the ontology

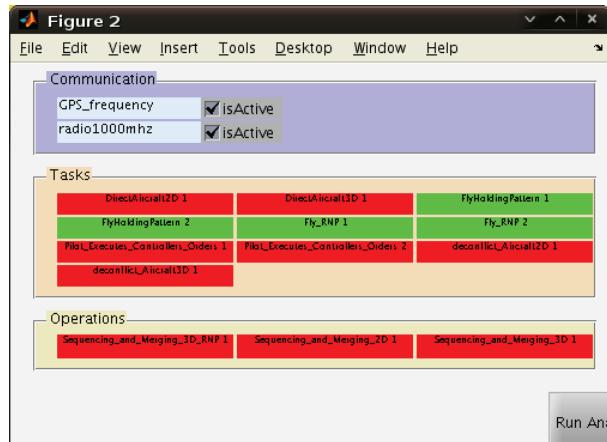


Fig. 6. MATLAB interface - Operations control panel for the ontology

V. CONCLUSION

This paper presented an ontology for the air traffic system based on a physical and a functional decomposition of the system. The physical decomposition includes facilities, aircraft, human operators and communication media and is linked to the functional decomposition of the operations into tasks and functions. Failures of systems or subsystems can be introduced at different level and their impact can be tracked all the way to the decrease of performances in the operations. As new technologies are introduced to leverage new concepts of operation, this ontology allows to study their modes of failure and find alternative solutions to ensure the graceful degradation of the ATS.

ACKNOWLEDGMENTS

This work was supported by Thales ATM inc, and under NASA Grant NNX08AY52A. The author thanks NASA Ames Research Center for the support offered to attend the conference. The authors would like to thank Daniel Muller from Thales ATM inc and Steve Bradford from FAA for their useful insights.

REFERENCES

- [1] Joint Planning and Development Office. NextGen. <http://www.jpdo.gov/>.
- [2] Eurocontrol, The European Organisation for the safety of Air Navigation. SESAR. The Single European Sky ATM Research Programme. <http://www.eurocontrol.int/sesar/>.
- [3] P. Kopardekar, K.D. Bilmoria, and B. Sridhar. Airspace configuration concepts for the next generation air transportation system. *Air Traffic Control Quarterly*, 16(4):313–336, 2008.
- [4] Wikipedia. US Airways Flight 1549, January 15 2009. http://en.wikipedia.org/wiki/US_Airways_Flight_1549.
- [5] CNN. Atlanta airport reopens after lightning strike, April 2009. <http://edition.cnn.com/2009/US/04/23/ga.airport.storms/index.html>.
- [6] pp prune.org. Southern California tracon (sct) evac, October 2003. <http://www.pprune.org/atc-issues/106897-southern-california-tracon-sct-evac.html>.
- [7] The Sweep / North County Times. Southern California tracon goes mapless, May 2007. <http://atcmuseum.wordpress.com/2007/05/25/southern-california-tracon-spends-a-mapless-hour/>.
- [8] Guardian.co.uk. Flight delays continue after air traffic control failure, September 2008. <http://www.guardian.co.uk/uk/2008/sep/26/transport.theairlineindustry>.
- [9] Encyclopedia.com. Air traffic control failure is examined, October 2007. <http://www.encyclopedia.com/doc/1Y1-111174919.html>.
- [10] Highbeam.com. Atc zero: even with backup systems and master plans, a catastrophic failure may still have critical air traffic control being done by cell phone.(system notes), December 2007. <http://www.highbeam.com/doc/1G1-203027937.html>.
- [11] Memphis Business Journal. Telecom glitch stops departures at memphis international, September 2007. <http://memphis.bizjournals.com/memphis/stories/2007/09/24/daily12.html>.
- [12] Yahoo News. Air traffic control failure is examined, October 2007. http://www.newsmgr.com/nm2/uploads/101107_ap_air_traffic_control_failure_is_examined.pdf.
- [13] A. Pinker and C. Smith. Vulnerability of the GPS Signal to Jamming. *GPS Solutions*, 3(2):19–27, 1999.
- [14] Airline Biz Blog. How the faa scolded the pilots, October 2009. <http://aviationblog.dallasnews.com/archives/2009/10/how-the-faa-scolded-the-pilots.html>.
- [15] T.B. Sheridan. Next Generation Air Transportation System: Human-Automation Interaction and Organizational Risks. In *Proceedings of the 2nd Symposium on Resilience Engineering*, November 8-10, 2006, Juan-les-Pins, France.
- [16] C. Nijssen, K. Eyferth, and T. Bierwagen. Modelling cognitive processes of experienced air traffic controllers. *Ergonomics*, 42(11):1507–1520, 1999.
- [17] S. M. Lee, S.Y. Kim, K. Feigh, and V. Volovoi. Structural framework for performance-based assessment of atm systems. In *9th AIAA Aviation Technology, Integration, and Operations Conference*, September 21 - 23, 2009, Hilton Head, South Carolina.
- [18] J. Daams, HAP Blom, and HB Nijhuis. Modelling human reliability in air traffic management. *FRONTIERS SCIENCE SERIES*, 2:1193–1200, 2000.
- [19] T. Prevot, E. Palmer, N. Smith, and T. Callantine. A multi-fidelity simulation environment for human-in-the-loop studies of distributed air ground traffic management. In *AIAA Modeling and Simulation Conference and Exhibit*, August 5 - 8, 2002, Monterey, CA.
- [20] S.T. Shorrock and B. Kirwan. Development and application of a human error identification tool for air traffic control. *Applied Ergonomics*, 33(4):319–336, 2002.
- [21] O. J. Pinon, K. Fry, and J.-P. Clarke. The air transportation system as a supply chain. In *AIAA Guidance, Navigation, and Control Conference and Exhibits*, August 10 - 13, 2009, Chicago, Illinois.
- [22] O. J. Pinon, E. Garcia, and D. Mavris. A methodological approach for airport technology evaluation and selection. In *26th International Congress of the Aeronautical Sciences*, September 14 - 19, 2008, Anchorage, Alaska.
- [23] T. Ritchey. General Morphological Analysis. *A general method for non-quantified modeling*, 16th EURO, 1998.
- [24] MD Di Benedetto, S. Di Gennaro, and A. D’Innocenzo. Error detection within a specific time horizon and application to air traffic management. In *44th IEEE Conference on Decision and Control, and European Control Conference*, pages 7472–7477, 2005.
- [25] M.R. Endsley and D.J. Garland. *Situation awareness: analysis and measurement*. CRC Press, 2000.
- [26] J.M. Sussman. *Perspectives on intelligent transportation systems (ITS)*. Springer, 2005.
- [27] R.S. Dodder, J.M. Sussman, and J.B. McConnell. The Concept of the CLIOS Process: Integrating the Study of Physical and Policy Systems Using Mexico City as an Example. In *Massachusetts Institute of Technology Engineering Systems Symposium, Cambridge, MA*, volume 31. Citeseer, 2004.
- [28] Federal Aviation Administration - Eurocontrol. Safety techniques and toolbox. Technical report, Version 2.0, October 3, 2007.
- [29] DF Haasl, NH Roberts, WE Vesely, and FF Goldberg. Fault tree handbook. Technical report, NUREG-0492, Nuclear Regulatory Commission, Washington, DC (USA). Office of Nuclear Regulatory Research, 1981.
- [30] NG Leveson, SS Cha, and TJ Shimeall. Safety verification of ada programs using software fault trees. *IEEE software*, 8(4):48–59, 1991.
- [31] M. Gariel and E. Feron. Graceful Degradation of Air Traffic Operations: Airspace Sensitivity to Degraded Surveillance Systems. *Proceedings of the IEEE*, 96(12):2028–2039, 2008.
- [32] M. Gariel, E. Feron, and J.P. Clarke. Air Traffic Management complexity maps induced by degradation of Communication, Navigation and Surveillance. In *AIAA Guidance, Navigation and Control Conference and Exhibit*, August 18 - 21, 2008, Honolulu, Hawaii.
- [33] M. Gariel and E. Feron. 3D Conflict Avoidance under Uncertainties. In *IEEE/AIAA 28th Digital Avionics Systems Conference*, pages 4.E.3–1 – 4.E.3–8, October 23–28, 2009, Orlando, Florida.
- [34] S.Y. Kim, K. Feigh, S.M. Lee, and E. N. Johnson. a task decomposition method for function allocation. In *AIAA Infotech@Aerospace Conference*, April 6 - 9, 2009, Seattle, Washington.
- [35] V. Volovoi, G. Calanni Fraccone, M. Hedrick, R. Kelley, and A. Colon. Agent-based simulation of off-nominal conditions during a spiral descent (nextgen vehicle nra). In *9th AIAA Aviation Technology, Integration, and Operations Conference*, 2009, Hilton Head, South Carolina.

Game Equilibrium Analysis on an Auctioning Method for Single Airport Congesting Resource Allocation

LIU Fangqin

College of Civil Aviation

Nanjing University of Aeronautics and Astronautics

Nan Jing, China

liuliuqin2003@yahoo.com.cn

HU Minghua

College of Civil Aviation

Nanjing University of Aeronautics and Astronautics

Nan Jing, China

minghuahu@263.net

Abstract—Flight delays cause lots of additional operational costs of airlines. Because airspace capacity is a scarce resource and airlines are self-interested, how to optimize the capacity allocation and avoid “the tragedy of the commons” is a hard problem for the Air Traffic Management authority. Through defining the marginal cost function and the opportunity cost function about the airlines, we introduce the first-price-sealed bid theory to realize the scarce capacity allocation of congesting airport to the airlines which want them. Under the ATM authority’s resource allocation policy, the airline will develop a set of scenarios to minimize the potential disruption to its schedule and implement the one that is most cost-effective through a competitive bidding process with other airlines. Finally, the Air Traffic Management authority could get the optimized global allocation result of airspace resources under the equilibrium condition.

Keywords- air traffic management; ground delay program; first pricing sealed auction; traffic congestion

I. INTRODUCTION

During the last several years, airlines and passengers have been suffering from congestion at busy airports and airspaces, annual congestion delay costs airlines and travelers more than \$20 billion in the world. Especially in china, with the greatly development of trades and economics in the last two decades, the air traffic have been growing rapidly and the flights delays have become a major public policy issue. It is predicted that there will be more than 3655000 flights operating in 2010. Though comparing with the European and USA, she has fewer than any of them, the distribution of flight flows in China is very uneven. More than 70% of flights have been operating on the eastern china. Nearly 30% of total flights have been flying between these three airports (including Beijing airport and Shanghai airport and Guangzhou airport) every year. and even the uneven trend is developing. So the Chinese air traffic operational system is fragile and is fluctuated easily. The air traffic congestion and delay problem is more and more prominent due to the military actions and bad weather conditions.

At present, traffic flow management(TFM) with the collaborative decision making aid(CDM) is applied widely to help to resolve the traffic congestion and balance demand and capacity when the airspace system was disrupted or the capacity was decreasing due to bad weathers or military events in the airspace system. A common condition in airline schedule planning (the process of generating the schedule with the greatest revenue potential) is that flight legs will be operated as planned according to the natural capacity of airspaces. But, when bad weather happened on the airspace, the capacity of airspace is decreased largely. The uncertain shortfall of capacity disrupts the planed schedules, and lots of affected flights need to be rescheduled. According to the schedule time, different traffic management initiatives such as reroutes, ground delay programs and miles-in-trail (MIT) restrictions can be used to revise the flight schedules and make the schedule demand adapt to the airspace system.

How to reschedule and allocate the limited airspace resources to these disrupted flights of different airlines equitably and efficiently under these three traffic management initiatives is the hard problem for the ATM authority. ATM authority aims to minimizing system delay time or cost under some certain fairness rules when it reschedules the disrupted flights of different users. Minimizing system delay time does not reflect the lowest total delay cost. Being incorporated into the collaborative decision making (CDM) process, the airlines could influence the rescheduling decisions to profit themselves [1][2]. Because the total delay cost does not include the airlines’ delay costs but also include the travelers’ delay costs, the goals of different decision makers which include airlines and ATM authority may conflict and the available information for good decision makings varies among these decision makers. The airlines maybe hide the flight information that is disadvantageous to them, but is necessary to the optimal system decision. It is hard for ATM authority to get to the aim of the decision that is to reschedule these disrupted flights and allocate the limited airspace resources to airlines equitably and efficiently.

The auction is a resource rationing method of the market mechanism. The bid price is the reflection of the value of a scarce resource for the bidder. The successful use of auctions

for telecommunication spectrum, energy and other commodities provide valuable insight into how to design auctions for the airspace resources [3-5]. Due to the fast progress of network and web technologies, traditional trading systems can be operated well on the internet. It unchained the technical barrier for the auction applied to the air traffic management [4] [6].

In this paper, we presents a first sealed auction method based on Dynamic Stackelberg equilibrium to realize the coincidence goal between the ATM authority and airlines. We make an attempt to set up the market-based, user self-decision Air Traffic Management mechanism. ATM authority sets up and announces the specific congestion toll schedules for the performance of the system that internalize the congesting external cost into the flight operational cost of airlines. ATM authority takes into consideration the global impact of dynamic congestion tolls that encourages the profit-oriented airlines to shift their low marginal profit flights to the non-peak traffic period or other legs which may be not charged by congesting fees or charged a little. Each airline is assumed to reschedule its disrupted flights according to the maximizing self-interested rule, while taking into consideration the pre-announced toll schedules and allocated capacity which is preferentially sold to the airline. Those elastic flights may be shifted from the congesting airspace to be delayed or to reroute other airspace.

II. THE AUCTION EQUILIBRIUM MODEL FOR THE CERTAIN CAPACITY

We first define the usage cost of airspace r and the expectation delay cost of flight f_k and the opportunity cost of flight f_k .

Let the marginal usage cost of the certain capacity of airspace r as follow,

$$MC_{a_i}^1(f_k, t) = p_{a_i}^1(t) \quad (1)$$

Where $p_{a_i}^1(t)$ denotes the bidding price of the certain capacity that airliner a_i submit for flight f_k

Let the lower one of the expectation delay cost and the rerouting cost of flight f_k as the opportunity cost of f_k using the certain capacity, as follow

$$OP_{f_k}(t) = \begin{cases} delay_{f_k}(t + \Delta t_{f_k}) - delay_{f_k}(t) \\ +(1 - prob_{gdp}(t)) * p_{a_i}(t + \Delta t_{f_k}), & \text{if } Edelay_{f_k}(t) < reroute_{f_k}(t), \\ reroute_{f_k}(t), & \text{if } reroute_{f_k}(t) < Edelay_{f_k}(t). \end{cases} \quad (2)$$

Where

$$\begin{aligned} Edelay_{f_k}(t) &= prob_{gdp}(t) * [delay_{f_k}(t + \Delta t_{f_k}) - delay_{f_k}(t)] \\ &\quad + \overline{prob_{gdp}(t)} * [MC_{a_i}^1(f_k, t + \Delta t_{f_k}) \\ &\quad + delay_{f_k}(t + \Delta t_{f_k}) - delay_{f_k}(t)] \\ &= delay_{f_k}(t + \Delta t_{f_k}) - delay_{f_k}(t) \\ &\quad + (1 - prob_{gdp}(t)) * p_{a_i}(t + \Delta t_{f_k}) \end{aligned} \quad (3)$$

Where

The expectation delay cost at the future departing interval $(t + \Delta t_{f_k})$, we define in this paper, is the operational cost after receiving the Ground Delay Program (GDP) Order, if the flight do not depart at the current interval. After a certain length $(t + \Delta t_{f_k})$ predicted by the airliner, if the congestion problem will not exist, the GDP initiative will be canceled and the delay cost will be the ground delay cost from the current interval. But, if the situation has not become good, the cost should include the ground delay cost and the usage cost of airspace r_j at the future departing interval $(t + \Delta t_{f_k})$. $prob_{gdp}(t)$ is the probability that the GDP initiative will be canceled during the $(t + \Delta t_{f_k})$ th period. So, the delay cost of flight f_k is the expectation value including the ground delay cost with probability $prob_{gdp}(t)$ and the usage cost at the future departing interval $(t + \Delta t_{f_k})$ with probability $1 - prob_{gdp}(t)$.

If the airliner wins the certain capacity bidding game, the payoff utility of flight f_k is as follow,

$$\begin{aligned} OP_{f_k}(t) - MC_{a_i}^1(f_k, t) \\ = \begin{cases} Edelay_{f_k}(t) - p_{a_i}^1(t), & \text{if } delay_{f_k}(t + \Delta t_{f_k}) - delay_{f_k}(t) < reroute_{f_k}(t), \\ reroute_{f_k}(t) - p_{a_i}^1(t), & \text{if } reroute_{f_k}(t) < delay_{f_k}(t + \Delta t_{f_k}) - delay_{f_k}(t). \end{cases} \\ \approx \begin{cases} delay_{f_k}(t + \Delta t_{f_k}) - delay_{f_k}(t) - prob_{gdp}(t) * p_{a_i}^1(t), & \text{if } delay_{f_k}(t + \Delta t_{f_k}) - delay_{f_k}(t) < reroute_{f_k}(t), \\ reroute_{f_k}(t) - p_{a_i}^1(t) & \text{if } reroute_{f_k}(t) < delay_{f_k}(t + \Delta t_{f_k}) - delay_{f_k}(t). \end{cases} \end{aligned} \quad (4)$$

Where

Because the unknown expectation delay cost is the empirical data, for simplifying the problem we take the delay cost of flight instead of the expectation delay cost to the rerouting cost.

Only if is the marginal usage cost of the certain capacity lower than its opportunity operational cost, the airline will attend the auction for flight f_k . So, the payment utility value is always a positive number. Simply, we assume that

$p_{a_i}(t + \Delta t_{f_k})$ is approximate to $p_{a_i}^1(t)$ if there will be still the congestion during the future departing interval $(t + \Delta t_{f_k})$. Both opportunity operational cost $OP_{f_k}(t)$ of flight f_k and marginal cost of the uncertain capacity are the additional operational cost of the flight. We assume that the additional operational cost, caused by congestion-related events, of each flight that takes part in the auction are independent and uniform random variables on the same interval $(0, delay_A(T))$. Because all of users in set A are profit-oriented, we assume that in civil aviation industry there is the common maximum of the additional operational cost of flight -- $delay_A(T)$. Each of the bidders who auction the same resource submits a nonnegative bidding price. The bidder submitting the highest bid price will win and pay his bid. Other bidders pay and receive nothing. Bidders are risk-neutral and all of this information is common knowledge. If bidder a_i wins and pays the bidding price, bidder a_i 's payoff is

$$bid_{a_i, f_k}(a_i, a_j, v_{a_i, f_k}(t)) = \begin{cases} OP_{f_k}(t) - p_{a_i}^1(t), & \text{if } p_{a_i}^1(t) > p_{a_j}^1(t), \\ 0, & \text{or else,} \end{cases}$$

$$\forall a_j \in A. \quad (5)$$

Because the bid game is peer to peer, we just need to analyzing the equilibrium strategy of $a_i : p_{a_i}^1(t) = p_{a_i}^{1*}(OP_{f_k}(t))$. Given the equilibrium solution $OP_{f_k}(t) - p_{a_i}^{1*}(t)$, the expected payoff function is as follow,

$$Ebid_{a_i, f_k}(p_{a_i}^1(t)) = (OP_{f_k}(t) - p_{a_i}^1(t)) * [\prod_{j \neq i} \text{probability}(p_{a_j}^1(t) < p_{a_i}^{1*}(t))] \quad (6)$$

Where

The first part before the multiplicative sign is the payoff of a_i , the second part is the probability that a_i wins all of the others.

The probability that bidder a_i wins bidder a_j is

$$\begin{aligned} & \text{probability}(p_{a_j}^1(t) < p_{a_i}^{1*}(t)) \\ &= \text{probability}(p_{a_j}^{1*}(OP_{f_k}(t)) < p_{a_i}^{1*}(t)) \\ &= \text{probability}(OP_{f_k}(t))_{a_j} < \Phi(p_{a_i}^{1*}(t)) \\ &= \Phi(p_{a_i}^{1*}(t)) / delay_A(T) \end{aligned} \quad (7)$$

Where $\Phi(p_{a_i}^{1*}(t))$ is the inverse function of $p_{a_i}^{1*}(t)$, which denotes additional operational cost saving is $\Phi(p_{a_i}^{1*}(t))$ if airline a_i submitted bid price $p_{a_i}^{1*}(t)$. So, we get

$$\begin{aligned} Ebid_{a_i, f_k}(p_{a_i}^{1*}(t)) &= (OP_{f_k}(t) - p_{a_i}^1(t)) \\ &\quad * [\Phi(p_{a_i}^{1*}(t)) / delay_A(T)]^{n-1} \end{aligned} \quad (8)$$

For maximizing the expected payoff, we get

$$\frac{\partial Ebid_{a_i, f_k}(p_{a_i}^{1*}(t))}{\partial p_{a_i}^{1*}(t)} = 0$$

If $delay_{f_k}(t + \Delta t_{f_k}) - delay_{f_k}(t) < reroute_{f_k}(t)$, we get

$$\begin{aligned} \frac{\partial Ebid_{a_i, f_k}(p_{a_i}^{1*}(t))}{\partial p_{a_i}^{1*}(t)} &= -prob_{gdp}(t) * [\Phi(p_{a_i}^{1*}(t))]^{n-1} \\ &\quad + [delay_{f_k}(t + \Delta t_{f_k}) - delay_{f_k}(t)] \\ &\quad - prob_{gdp}(t) * p_{a_i}^1(t)] * (n-1)\Phi^{n-2}\Phi'(p_{a_i}^{1*}(t)) \\ &= 0. \end{aligned}$$

If $reroute_{f_k}(t) < delay_{f_k}(t + \Delta t_{f_k}) - delay_{f_k}(t)$, we get:

$$\begin{aligned} \frac{\partial Ebid_{a_i, f_k}(p_{a_i}^{1*}(t))}{\partial p_{a_i}^{1*}(t)} &= -[\Phi(p_{a_i}^{1*}(t))]^{n-1} + [reroute_{f_k}^*(t) \\ &\quad - p_{a_i}^{1*}(t)] * (n-1)\Phi^{n-2}\Phi'(p_{a_i}^{1*}(t)) \\ &= 0 \end{aligned}$$

Due to $\Phi(p_{a_i}^{1*}(t)) = \min(OP_{f_k}(t), reroute_{f_k}(t))$, we get the equilibrium bid price of a_i , as follow,

$$p_{a_i}^{1*}(t) = \begin{cases} \frac{n-1}{n * prob_{gdp}(t)} [delay_{f_k}(t + \Delta t_{f_k}) - delay_{f_k}(t)] \\ - \frac{1}{n} [delay_{f_k}(t + \Delta t_{f_k}) - delay_{f_k}(t)]^{1-n}, \\ \quad \text{if } delay_{f_k}(t + \Delta t_{f_k}) - delay_{f_k}(t) < reroute_{f_k}(t), \\ \frac{n-1}{n} reroute_{f_k}(t), \\ \quad \text{if } reroute_{f_k}(t) < delay_{f_k}(t + \Delta t_{f_k}) - delay_{f_k}(t). \end{cases} \quad (9)$$

Where the equilibrium price of this bid game relies on the number of bidders and the value $v_{a_i, f_k}^*(t)$ and their own estimations about the GDP delay situation. Each airline's bidding price is determined by the value from the bidding resource. At the equilibrium condition, the airline who gets the highest value from the resource will give the highest price. According to the first price sealed bidding principle, the player who gives the highest price will get the resource.

We get the differential equation of $p_{a_i}^{1*}(t)$ about the derivative $v_{a_i, f_k}(t) = delay_{f_k}(t + \Delta t_{f_k}) - delay_{f_k}(t)$,

$$\frac{\partial p_{a_i}^{1*}(t)}{\partial v_{a_i, f_k}(t)} = \frac{n-1}{n * prob_{gdp}(t)} - \frac{1-n}{n} [v_{a_i, f_k}(t)]^{-n} \geq 0, \quad (10)$$

Obviously, $p_{a_i}^{1*}(t)$ is the increasing function about the variable $v_{a_i, f_k}(t)$. So the airlines whose flights suffer the more

delay or rerouting cost will give the more prices about the auctioned resources, and will get more chance of winning.

When the bidding resources are more than one, if the bidding flights has the consistent utility for each resource unit in the same decision period, based on (1)–(5), similarly, we could get,

$$\begin{aligned} Ebid_{a_i, f_k}(p_{a_i}^{1*}(t)) = \\ (OP_{f_k}(t) - p_{a_i}^1(t)) * [\Phi(p_{a_i}^{1*}(t)) / delay_A(T)]^{n-m} \end{aligned} \quad (11)$$

Due to there being m available resources, if only the bidding price of the flight is above to any of $(n-m)$ other bidders, not to any of $(n-1)$ other bidders, the flight could win one capacity unit.

Likewise in (9), we get the equilibrium bidding price under the condition that the bidding resources are more than one, as follow,

$$\begin{aligned} p_{f_k, a_i}^{1*}(t) = \\ \left\{ \begin{array}{l} \frac{n-(m-1)-1}{n-(m-1)*prob_{gdp}(t)} [delay_{f_k}(t + \Delta t_{f_k}) - delay_{f_k}(t)] \\ - \frac{1}{n-(m-1)} [delay_{f_k}(t + \Delta t_{f_k}) - delay_{f_k}(t)]^{1-n+(m-1)}, \\ \text{if } delay_{f_k}(t + \Delta t_{f_k}) - delay_{f_k}(t) < reroute_{f_k}(t), \\ \frac{n-(m-1)-1}{n-(m-1)} reroute_{f_k}(t), \\ \text{if } reroute_{f_k}(t) < delay_{f_k}(t + \Delta t_{f_k}) - delay_{f_k}(t). \end{array} \right. \end{aligned} \quad (12)$$

Where

Let m denotes the auctioned certain capacity number.

Here the equilibrium value is just theory results. In practice, the behaviors of airlines in the bid games are hard to be assumed. However, the big and small of the equilibrium bid price is direct correlative to the additional operational cost. The

equilibrium bid price of flight could reflect the true additional operational cost of flight in the assumption that every airline is rational and profit-oriented.

III. CONCLUSIONS

Given that the additional operational costs of flights are important components of the airline decision-making process, how the economic costs of flights under the different air traffic management tactics influence the airline decision behaviors have not analyzed in previous research. Our models allow for a test of the market mechanism effects on the airline decision behaviors in the context of air traffic management (ATM) that carefully optimizes the airspace system costly. The main contribution of this paper is to develop the auction method of the market mechanism. In theory the first pricing sealed auction could ensure the systemic benefit and equity and efficiency. The method makes an attempt to solve the airport congesting capacity allocation problem in the pre-tactics air traffic flow management.

REFERENCES

- [1] L.Wojcik, "Models to understand airline and air traffic management authority decision-making interactions in schedule disruptions: From simple games to agent-based models", Handbook of Airline Strategy, 2001, pp. 549-575.
- [2] L.A.Wojcik, "Three principles of decision-making interactions in traffic flow management operations", Proc. 4th USA/EUROPE Air Traffic Management R&D Seminar Santa Fe, NM, USA 2001.
- [3] M. Ball, G. Donohue and K. Hoffman, "Auctions for the safe, efficient and equitable allocation of airspace system resources", Combinatorial Auctions, 2006, vol.1.
- [4] M. K. Franklin and M. K. Reiter, "The design and implementation of a secure auction service," IEEE Transactions on Software Engineering vol.5, no.22, 1996, pp. 302-312.
- [5] W.S. JUANG, H.T. LIAW, P.C. LIN and C.K. LIN, "The design of a secure and fair sealed-bid auction service", Mathematical and Computer Modelling vol.41, 2005, pp. 973-985.
- [6] S. Subramanian, "Design and verification of a secure electronic auction protocol", in Proc. of IEEE 17th ta Symposium on Reliable Distributed Systems, 1998.

Fair Slot Allocation of Airspace Resources Based on Dual Values for Slots

Nasim Vakili Pourtaklo

School of Electrical and Computer Engineering
and Institute for Systems Research
University of Maryland
College Park, Maryland 20742
Email: nasimv@umd.edu

Michael Ball

Robert H Smith School of Business
and Institute for Systems Research
University of Maryland
College Park, Maryland 20742
Email: mball@umd.edu

Abstract—Fair allocation of available resources among airlines is very challenging when there is a reduction in en-route resources. Each airline will typically place a different relative weight on delays, rerouting and cancelation. Whereas some airlines would like to preserve the on-time performance for certain flights and cancel or reroute many other flights, other airlines prefer to have less rerouting and cancelations while tolerating higher total delay. The value (or cost of delay) an airline associates with a particular flight may vary substantially from flight to flight. Airlines who wish to receive priority for certain flights usually are willing to pay more for specific time slots. To accommodate richer carrier preferences so that airlines can express the relative importance of delays, rerouting and cancelations, new concepts of slot values and dual pricing are introduced in this research. Unlike Ration By Schedule (RBS), the current algorithm in use for rationing airspace resources, that gives priority based on scheduled flight arrival times, our new allocation method provides flexibility to carriers to achieve their goals. Specifically, it also allows carriers to receive “premium” slots for an extra “charge”. In this paper, we describe a new rationing and randomized allocation method. We analyze the performance of the new method and compare it with RBS based on data derived from a real application. Our method has potential usefulness both in Airspace Flow Program (AFP) planning and in the emerging System Enhancements for Versatile Electronic Negotiation (SEVEN).

Keywords-resource rationing; flow management; fairness; equitable allocation; AFP; Dual Price

I. INTRODUCTION

When there is a capacity reduction due to the severe weather, rerouting flights is not sufficient to address extended capacity reductions in the airspace, and the need for additional tools has long been recognized. To meet that need the FAA (Federal Aviation Administration) introduced a new capability in the spring of 2006. The Airspace Flow Program (AFP) combines the power of Ground Delay Program (GDP) and Flow Constraint Area (FCA) to allow more efficient, effective, equitable, and predictable management of airborne traffic in congested airspace.

When TFM specialists at the Air Traffic Control System Command Center (ATCSCC), in consultation with FAA field managers and customer representatives, decide that the

weather conditions are appropriate they can plan and deploy an AFP. The first step is to use the Traffic Situation Display (TSD) to examine predicted weather and traffic patterns and identify the problem area by creating an FCA. An FCA is a user-defined volume of airspace along with associated flight lists and filters. FCAs are used to show areas where the traffic flow should be evaluated or where initiatives should be taken due to severe weather or volume constraints. Traffic managers or flight dispatchers define a geographic area of an FCA by drawing a polygon or a line on the display and defining the ceiling and floor of the FCA using a dialog box. FCAs are built by the ATCSCC and require a traffic management initiative (TMI).

The Enhanced Traffic Management System (ETMS) takes the FCA description and produces a list of the flights that are expected to pass through the FCA and the time they are expected to enter. This list, updated with fresh information every five minutes, is sent to the Flight Schedule Monitor (FSM), which displays the projected demand in a number of formats designed to support effective planning. FSM creates a common situational awareness among all users and service providers in the National Airspace System. All parties need to be aware of NAS constraints in order to make collaborative air traffic decisions. It is designed to effectively interact with existing FAA systems, FSM displays the Aggregate Demand List (ADL) information for both airport and airspace data elements for its users, which means everyone is looking at the same picture. The TFM specialists at the ATCSCC can enter the capacity of the FCA, expressed as the number of flights that can be managed per hour, and FSM will then assign each flight a controlled departure time so that the flow into the FCA does not exceed the declared capacity. These departure times are sent to the customers for flight planning and to the towers at the departure airports for enforcement.

The principal goal for the initial deployment of the AFP program is to better manage en route traffic during severe weather events. Compared to previous approaches, AFP's reduce unnecessary delays while providing better control of demand, more equity, and more flexibility for customers [3].

Today, AFP's use GDP-like tools. However, there are im-

portant differences between resource allocation for GDP's and enroute resource allocation. First, a GDP only applies delays to a subset of flights destined for a single airport while the AFP's apply delays to a subset of flights predicted to fly through a designated FCA (GDP tools have been modified for AFP's in this respect). Second, in the GDP setting, demand is established based on the set of flights scheduled to arrive at the GDP airport; GDP procedures implicitly assume all flights must be assigned an arrival slot. On the other hand, in the AFP or SEVEN (System Enhancements for Versatile Electronic Negotiation) setting, all flights on the demand list need not be granted access to the enroute resource. The flight operator has the prerogative to cancel flights not given access *or reroute such flights around the restricted airspace*. Thus, enroute resource allocation decision models should both determine which flights gain access and assign an access time (slot) for those flights that do gain access. The last related difference is the existence of a fixed flight schedule on which to base resource allocation for GDP's. Ration-by-schedule uses, in a very fundamental way, the flight schedule as the basis for resource allocation. In concept this can be done for enroute problems by simply taking the schedule associated with the list of flights whose flight plans have been filed through the impacted enroute resource. A key difference, however, is that the filing a flight plan is a short-term action and, as a result, the possibility of flight operators trying to "game the system", e.g. by filing unnatural flight plans, is a very real possibility.

In practice, each airline will typically place a different relative weight on delays, rerouting and cancelation. Whereas some airlines would like to preserve the on-time performance for certain flights and cancel or reroute other flights, other airlines prefer to have less rerouting and cancelation while tolerating higher total delay.

Using fairness principles as a basis for allocating scarce resources provides our research with a novel focus. In fact, some proposals address the rationing of airport arrival capacity in the long run. Using methods ranging from auctions [9] and congestion pricing [7] to bargaining schemes [1]. The allocation of slots under Collaborative Decision Making (CDM) is different, in that slots must be assigned on a daily basis due to fluctuation in airport or en-route capacity. The dynamic nature of the allocation process makes it more complicated and fairness plays an important role in this environment.

The method we propose applies to a general class of airspace resource allocation problems and, in fact, we have designed it to also be applicable to the emerging SEVEN [2]. While SEVEN should potentially have a broad range of application contexts, the key feature that it brings to bear, which is not present in AFP's, is the ability for flight operators to express preferences among various options for the disposition of an individual flight. The ability for flight operators to express preferences is also a key feature of our proposed resource allocation method.

In this paper, we propose a new method for assigning AFP slots to flights and flight operators, which is fundamentally

different from the method currently used for GDP's and AFP's, ration-by-schedule (RBS). Our work uses as a starting point research on GDP's [14] and the investigation of RBS as a basis for fair resource allocation [13].

II. PROBLEM DESCRIPTION AND OVERVIEW OF PROCEDURE

In our research we assume that flights pass the boundary of FCA one at a time (this is consistent with current practice). Therefore we can express the capacity as the number of available time slots. We consider those flights that are "scheduled" to arrive at the boundary of FCA. Such a flight schedule can be derived based on each flights scheduled departure time and filed flight plan. Employing such a schedule can be problematic as it is not immune to gaming or strategic behavior on the part of flight operators.

The simple FCA capacity model employed allows the FCA to be characterized by a set of entry slots. Let $\mathcal{S} = \{s_1, s_2, \dots, s_m\}$ be the set of available slots and $\mathcal{F} = \{f_1, f_2, \dots, f_n\}$ be the set of flights. However, in general $n > m$, i.e. the number of flights is greater than the number of slots during the AFP. The capacity, c_j , of each slot s_j is considered one which means each slot can be used by a single flight. Suppose there are K carriers $\mathcal{A} = \{A_1, A_2, \dots, A_K\}$, and \mathcal{F}_i is the set of flights of carrier A_i . a_f is the time flight f is scheduled to arrive at the boundary of FCA and t_j is the time of slot s_j . Flight f can be assigned to any slots s_j with $t_j \geq a_f$. As with GDP planning, although flights are assigned to slots, we view the flight-to-slot assignment as a slot-to-flight assignment operator.

We break the process down into two steps:

- Step 1a:Determine a fair share, FS_i for each flight operator, A_i .
- Step 1b:Obtain flight operator flight-slot priority lists.
- Step 2: Allocate flights to slots in a manner consistent with the fair share determined in Step 1a and their flight priorities obtained in step 1b.

The fair share for each carrier can be found in many different ways. A principal goal we seek is to provide equity among carriers. The allocation of homogeneous demands, when the total demand exceeds total available resources is addressed in [8], [15], [16] and, in the case of scheduling problems, is treated in [6], [4], [5], [12] (these models correspond to the situation in which all flights arrive at the beginning of the AFP). Vossen [12] uses a heterogeneous demand model to treat the different arrival times of flights. To allocate slots to flights, he uses "proportional random assignment" which randomly assigns slots to the carriers in proportion to the number of a carrier's flights that can use a slot. In his method, slots sequentially are assigned to the carriers. The proportional random assignment method is a randomized allocation method. It is also time dependent. In the "proportional random allocation" method proposed by Moulin [4] there is no time dependency, which means that all agents can participate in the lottery at each time until their demand is met. In proportional random

assignment, agents participate in the lottery as long as they can use the slot under consideration. We will use this method as a way to determine a fair share to each flight [11] (and consequently each flight operator). However, we will *not* use it to actually allocate slots to flights.

The flight-to-slot assignment carried out in Step 2 is a type of randomized round-robin that employs flight-operator preferences. In step 1b, each flight operator specifies an ordered list of flight-to-slot assignments. The allocation procedure gives priority to the carriers who wish to maintain their on-time performance for certain key flights and in return receive fewer slots. At each iteration, when a flight operator has its “turn”, the highest available assignment on that flight operator’s preference list is chosen. Here, by available, we mean the the associated flight has not yet been assigned a slot and the associated slot has not been assigned to a flight.

In Section III, we describe the procedure for determining FS_i , i.e. Step 1a and also explain the submission of flight priority list by each flight operators, Step 1b. We note that these procedures were previously described in [11] so this section is largely a review. Section IV covers Step 2, which is a new contribution. Section V provides our experimental results.

III. FAIR SHARE AND FLIGHT PRIORITY LIST

A. Determining Fair Share of Each Carrier From Available Slots

As discussed above the goal of this section is to determine a fair share of available slots “owed” to each operator in expectation. Our procedure for determining this fair share requires as input a flight schedule. Vakili and Ball [11] explained how to determine the fair share of available slots. In this section, we just briefly explain the procedure, *Finding Fair Share based on Proportional Random Assignment*, FFS-PRA.

The availability of a schedule is characterized by knowing for each flight f , a scheduled arrival time a_f , which is interpreted as the time f is “scheduled” to arrive at the FCA boundary. Each slot s_j has an associated time t_j so that a flight f can be assigned to slot s_j if $a_f \leq t_j$.

We start by assuming there is an allocation that uses all slots (this almost always happens during congested periods – furthermore, this assumption can be dropped but doing so would complicate the presentation). We call an allocation that uses all slots a *complete* allocation. PRA, which underlies FFSPRA is based on the following principles.

- Each flight can use *at most* one slot.
- All flights have equal share of each slot that they can use in any complete allocation.
- Each flight can be assigned to any slot later than its scheduled time of arrival.

We can now define the PRA procedure.

PRA:

- Step 1** : Set $F_1 = \{f \in \mathcal{F} : a_f \leq t_1\}$ and $i = 1$
Step 2 : Choose an $f \in F_i$ with probability $\frac{1}{|F_i|}$ and assign f to s_i

Step 3 : Set $i = i + 1$

Step 4 : Set $F_i = \{f \in \mathcal{F} : a_f \leq t_i\} - \{f\}$

Step 5 : If $i \leq m$ Then go to Step 2.

End.

We define for each flight f and slot j , P_{fj} to be the probability that PRA assigns f to s_j . Also, define:

$$P_f = \sum_j P_{fj} = \text{PRA share for flight } f \quad (1)$$

$$FS_i = \sum_{f \in \mathcal{F}_i} P_f = \text{PRA share for flight operator } A_i \quad (2)$$

Because of the structure of PRA, P_{fs} can be computed in polynomial time [11] as:

$$P_{fj} = \frac{\prod_{i=k}^{j-1} (n_i - i)}{\prod_{i=k}^j (n_i - (i - 1))} \quad (3)$$

where n_i is the number of flights that can be assigned to slot s_i , k is the earliest slot, s_k , that flight f can use.

Let us now compare this method of computing fare shares with the implicit fare shares allocated by RBS. RBS, of course, is a deterministic procedure that either assigns a slot to a flight or does not. Thus, the RBS “fare share” for a flight is either zero or one. Since PRA employs randomization and since it employs the principles described earlier, any flight that appears in *any* complete allocation will have a positive fare share. Therefore, all flights included in an AFP will have a positive share of available slots.

This is a very important point. While RBS will give zero share to later flights, FFS-PRA will give such flights a positive share. We should note that flights that are scheduled earlier will typically receive a higher share than later scheduled flights. Therefore, FFS-PRA implicitly gives a higher share to earlier scheduled flights and so it gives some weight to the basic RBS principle. However, it balances this principle with the principle that each flight included in the AFP has a claim to a portion of the available capacity.

Further, (see [10]) we can show that PRA meets the fundamental fair allocation principles, which are *impartiality*, *consistency* and *equal treatment of equals* and *demand monotonicity* (see [15]).

Impartiality states that allocation rule should not discriminate among the flights *except* insofar as they differ in type. In other words, if two flights are indifferent in type and in the feasible set, they will receive the same fair share. The consistency property states that the expected fair shares should be independent of the order in which flights are assigned to the slots. *Equal treatment of equals* states that if two flight operators have the same schedule, they will receive the same fair share. *Demand monotonicity* says that an increase in carrier i ’s total number of flights (with other flights remaining unchanged), can not deteriorate carrier i ’s fair share.

B. Flight Priority List

As discussed earlier our slot allocation procedure requires airline flight-slot preference information. There are two types of preference lists.

In the first type, carriers submit to the FAA an ordered list of flight-to-slot assignments. For example, carriers submit an ordered list of (f_i, s_j) pairs. This type of list can be very long when the number of slots is large. The second type of list is a compact version of the first type. Instead of submitting an ordered list of (f_i, s_j) pairs separately, carriers submit the pair of flights and an interval of slots. For example, if a carrier ordered preference list is $(f_i, s_j), (f_i, s_{j+1}), (f_i, s_{j+2}) (f_l, s_k)$ then it can be expressed as $(f_i, s_j : s_{j+2}), (f_l, s_k)$.

Suppose, carrier A has three flights A101, A102 and A103 and also assume there are six available slots, s_1, \dots, s_6 . The earliest slots, a_f , that each flight can be assigned could be:

Slot:	s_1	s_2	s_3	s_4	s_5	s_6
Flights:	A101	A102	A103			
a_f	s_1	s_4	s_6			

The following table illustrates a possible flight priority list.

Preference List for A

Rank	(Flight,Slot)	Rank	(Flight,Slot)
1	(A103, s_6)	6	(A101, s_3)
2	(A101, s_1)	7	(A101, s_4)
3	(A101, s_2)	8	(A101, s_5)
4	(A102, s_4)	9	(A101, s_6)
5	(A102, s_5)	10	(A102, s_6)

For simplicity, the flight priority list can be shown as:

Rank	(Flight,Slot)	Rank	(Flight,Slot)
1	(A103, s_6)	4	(A101, $s_3 : s_6$)
2	(A101, $s_1 : s_2$)	5	(A101, $s_5 : s_6$)
3	(A102, $s_4 : s_5$)	6	(A102, s_6)

In this example, for carrier A , the highest priority is to insure that flight A103 leaves on time. Thus, Carrier A would prefer to receive slot s_6 , before several earlier slots, in order to insure the ontime departure for flight A103.

IV. SLOT ALLOCATION PROCEDURE

Vakili and Ball [11] proposed a randomized allocation procedure, PBPRA, that uses fair share and flight priorities as exogenous input to allocate slots to flight operators. PBPRA guarantees that each carrier receives between the ceiling and floor of its fair share. However, in PBPRA, it was implicitly assumed that all slots had equal values. Specifically, when measuring an allocation against a carrier's fair share, they only considered the total number of slots a carrier received. The priority list allows carriers to express a preference among slots, however, it does not allow a carrier to trade off the number of slots received with which slots are received. For example, carriers that would like to maintain on time performance for key flights, may be willing to pay more than others for particular slots. We wish to allow carriers to "pay more" for earlier slots when they wish to do so.

Our objective here is somehow to distinguish between those carriers who want to maintain on-time performance for certain

flights and in return receive fewer slots and those carriers who can tolerate more delay but would like to receive more slots.

Consider the example of carrier A who prefers to receive priority for certain flights in exchange for receiving fewer slots in total. The algorithm employs a parameter which is the "value" of the higher priority slots distributed. Suppose that value was set at 2 "slot units" and that carrier A 's fair share is 5.5. Then carrier A could receive two "high-priority" slots based on 2 ($\lfloor \frac{5.5}{2} \rfloor$). The remainder of its fair share is 1.5, which can then be used to receive later slots. It is very important to notice that only those carriers that can afford this trade off (have a fair share ≥ 2) are considered). If a small carrier with a small fair share prefers to receive good slots and it does not have enough budget to give up a second flight, it can not be considered.

A. Slot Values

For illustration purposes, suppose we have two sets of airlines. Let A_1 be the set of airlines that prefer less delay and A_2 the set of airlines that prefer to receive more slots. In our allocation algorithm we initially give priority to the airlines in A_1 . Therefore, they must pay more for each slot they initially receive because of the priority. Let us assume the price of each slot they receive is P_H . Since airlines in A_1 receive priority in the allocation process their exogenous fair share must be greater than P_H .

The FAA acts as an independent, fair moderator. The FAA announces the value of priority slots. This value must be greater than one. The process operates so that the total value of slots given away equals the number of slots available. Since the value of each slot for the airlines in set A_1 is P_H , we can compute the value of remaining slots. Thus, later (less preferred) slots will have a value less than one. Suppose there are m slots available, to compute the value of remaining slots, we need to find the number of slots that are assigned to airlines in A_1 . Let us call this number m_1 :

$$m_1 = (\sum_{a \in A_1} [FS_a / P_H]) / P_H \quad (4)$$

Where FS_a is the fair share of carrier a . Then the value of remaining slots can be computed as:

$$P_L = \frac{m - P_H \times m_1}{m - m_1} \quad (5)$$

As we can see the value of the remaining slots is less than one. Note that higher P_H values result in a smaller m_1 . We will show the effect of varying P_H in our simulation results.

B. Dual Price Proportional Random Allocation

Dual Price Proportional Random Assignment (DP-PRA) is a new algorithm that considers the carriers' tradeoff between delay and rerouting (or cancellation). DP-PRA contains two phases: First phase allocates slots to the flights in the set A_1 and in the second phase all remaining slots are allocated from the earliest available to the latest available. The second phase can use the PBPRA [11].

We define two policies: under *Policy P₁* carriers prefer to prioritize certain flights, i.e. receive fewer slots but less delay; under *Policy P₂* carriers wish to treat all flight equally (and receive more slots). We use the notation $P_1 \succ_a P_2$ when carrier a prefers policy P_1 to policy P_2 .

We formally define DP-PRA below:

DP-PRA:

Step 0a: Inputs: Set of flights \mathcal{F} , set of carriers \mathcal{A} , set of available slots \mathcal{S} , Carriers' preference lists: $PList_1, PList_2, \dots, PList_K$ also P_H and carriers set $A_1 = \{a \in \mathcal{A} : P_1 \succ_a P_2, FS_a \geq P_H\}$.

Step 0b: Calculate the fair share of each airline FS_a based on PRA

Step 0c: Calculate P_L based on 4 and 5

Step 1: PHASE 1 while $A_1 \neq \emptyset$ Do:

Step 1a: $\forall a \in A_1$, Randomly choose an $a^* \in A_1$ in proportion to FS_{a^*} .

Step 1b: From $PList_{a^*}$, assign the best slot available to the highest priority flight (f^*, s^*)

Step 1c: $FS_{a^*} = FS_{a^*} - P_H$, $PList_{a^*} = PList_{a^*} - \{f^*\}$ and $\mathcal{S} = \mathcal{S} - \{s^*\}$ and $A_1 = \{a \in A_1 : FS_a \geq P_H\}$.

end while

Step 2: PHASE 2

Step 2a: $A = \{a \in \mathcal{A} : FS_a > 0\}$.

Step 2b: for all a in A , $FS_a = FS_a / P_L$.

Step 2c: Run PBPR.

In the first phase of algorithm we consider just carriers in A_1 who can afford a slot with value of P_H . A carrier will be chosen randomly based on its fair share, FS_{a^*} . Then from $PList_{a^*}$ we assign the best slot available to the highest priority flight, f^* . Assign f^* to s^* then remove f^* from $PList_{a^*}$ and s^* from \mathcal{S} . We reduce the fair share of a^* by P_H . We repeat this phase until A_1 becomes empty. Now, we move to the second phase.

In the second phase of the algorithm all airlines with positive fair share will be considered. The value of each slot in the second phase is P_L . We make the value of each slot one and increase the fair share of all airlines by $1/P_L$. Then, we execute PBPR. A carrier will be chosen randomly in proportion to its fair share. From $PList_a$ the highest priority flight from carrier a will be chosen. Carrier a 's fair share will be reduced by one.

We can show that DP-PRA satisfies some desirable properties:

- 1) **The value of slots allocated to a carrier A_i should be close to FS_i .** Let us consider the fair share as an exogenous budget each carrier has. This property, which is a version of "equal treatment of equals" is probably the most fundamental to consider. It says that each flight operator should get its fare share (within a tolerance). The actual total slot value for a carrier after using the procedure can be calculated based on whether slot value is P_H or P_L , and based on actual total number of slots

received by that carrier. After applying DP-PRA, then for any two carriers with equal fair share the difference in actual total slot value for two carriers with the same fair share will be less than an upper bound of $2P_L$. To be more precise, if two carriers with equal fair share belong to the same set, then the difference in actual total slot value for each carrier is less than P_L . And if two carriers with equal fair share belong to two different sets, then the difference in actual total slot value for each carrier is less than $2P_L$.

- 2) **Each flight operator should be motivated to submit a "truthful" preference list.** This is considered a fundamental property of allocation methods, more formally known as *strategy proofness*. If the "dominant" strategy for each flight operator is to submit its true priority list, then flight operators need not seek to "game the system" and so the problem they face is relatively straight forward. Further, the overall system will be more stable in the sense that there should not be claim that certain operators gained an unfair advantage. It can be shown that DP-PRA is strategy-proof if in each step of allocation procedure there are more than one carrier to compete for a slot.

V. EXPERIMENTAL RESULTS

In our experiment, we use the same test data set as used in [11]. This data set that had been employed by the CDM Future Concepts Team to perform human-in-the-loop experiments related to SEVEN. It contained 386 flights with 38 flight operators. The data included scheduled arrival arrival times at an FCA boundary. The FCA duration was from 18:00 pm to 21:00 pm. As we explained in [11] a flight cost function can be generated as:

$$C(x, P) = \begin{cases} 0 & x \leq 15 \\ (32 + 0.1P)(x - 15) & 15 < x \leq M_p \\ (32 + 0.1P)(M_p - 15) & x > M_p \end{cases}$$

Where M_p is flight specific max delay. Given the cost function, we generated the priority list for each flight operator based on all available flights that could use a slot; and the assumption is that the flight operator preferred allocating the slot to the flight with the highest marginal cost of delay. The flight operators are randomly assigned to set A_1 , the ones who prefer to receive better slots, or $A - A_1$, flight operators who prefer to receive more slots.

We compared the results of DP-PRA against ration-by-schedule (RBS), which is currently used to allocate FCA access during airspace flow programs. In our experiment, we considered 40%, 50%, 60%, 70% and 80% en-route capacity reduction for the FCA. We performed 2000 repetitions of the procedure since both procedures are random. In first part of all of our experiment we set $P_H = 2$. We will show later the effect of changing P_H . For each capacity reduction, the number of carriers that can participate in the first phase of algorithm is different. It is clear that as capacity increases the fair share

TABLE I
LIST OF AIRLINES THAT CAN PARTICIPATE IN THE FIRST PHASE AND THE NUMBER OF SLOTS ARE ASSIGNED

% Capacity reduction	List of Airlines	Number of slots
40	{1,3,5,6,21,25,28,29,34}	45
50	{1,5,6,21,25,28,29,34}	37
60	{1,5,6,21,25,29,34}	27
70	{1,5,6,21,25,29,34}	21
80	{1,21,29}	12

of each airline increases, consequently the number of airlines that can participate will increase as well. Airlines 1, 3, 5, 6, 7, 9, 17, 19, 20, 21, 25, 26, 28, 29, 30, 31, 34, 35 have the second policy. Table I shows the airlines and number of flights (or slots) that are assigned in the first phase for each capacity reduction. Table II shows the percentage of cost savings for

TABLE II
COMPARISON OF COST REDUCTION FOR DP-PRA AND MDP-PRA vs.
RBS

% Capacity reduction	DP-PRA
40	18.19
50	15.72
60	11.69
70	9.71
80	6.78

DP-PRA compared to RBS.

The main advantage of DP-PRA compared to PBPRA is to meet carriers' preference better. Figure 1 shows the average number of slots carriers in A_1 receives compare to previous procedures for 60% capacity reduction.

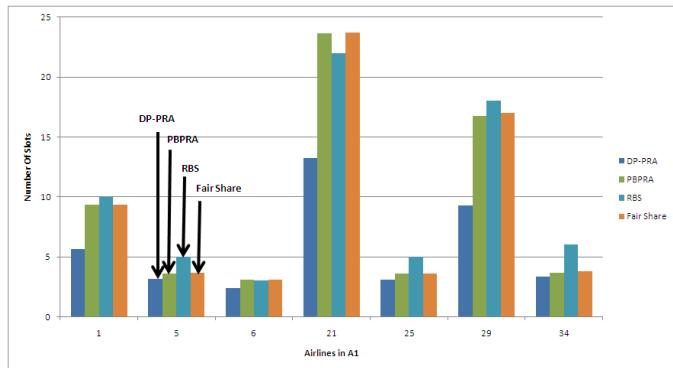


Fig. 1. Comparison of number of slots received for airlines in A_1

A. Effect of P_H

We have used $P_H = 2$ in our experiments. Here we want to investigate the effect of P_H in overall performance of DP-PRA. Choosing the right P_H is a challenge for the FAA.

There can be many different performance criteria; for example, deviation from carriers' fair share, total internal cost, how many slots should be assigned in the first phase. Minimizing the total internal cost is very hard for the FAA to measure because each carrier's cost information is private. Here we explain the effect of P_H on one performance criteria. In our examples we consider 40% capacity reduction in enroute resources.

The FAA can consider the deviation of total slot value received from fair share as a one criteria. Figure 2 shows the total define Minimum Square Error (MSE) of slot values from carriers' fair share. As can be seen, a minimum occurs at $P_H = 2.75$ and $P_H = 3.5$ for the procedure.

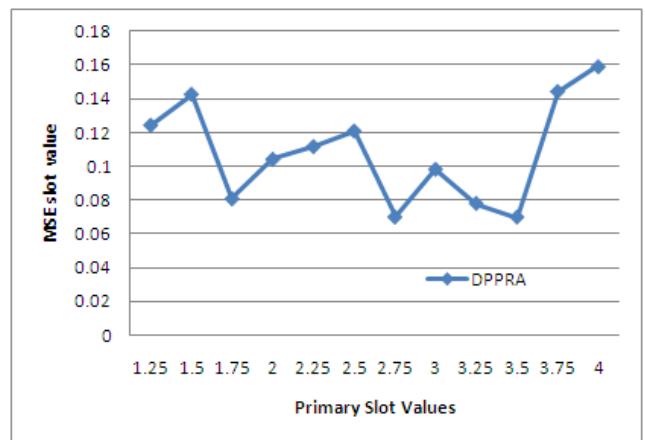


Fig. 2. Effect of primary slot values on MSE of slot values.

VI. CONCLUSIONS

In this paper a new procedure for slot allocation has been proposed. Unlike PBPRA that implicitly assigns the same value to all slots, under DP-PRA, we allow carriers to "pay more" to receive high priority slots. The main goal is to address carriers' preferences better. Our experiments show that, when using DP-PRA, carriers can better optimize their internal cost functions. Our procedure meets certain fair allocation principles, including equal treatment of equals and strategy proofness. A related challenge is to set (exogenously) the "price" for the high value slots. We have provided experiments to lend insight into this decision.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Robert Hoffman for providing us with flight data and also his help for computing cost function for the flight operators. This work was supported by the National Center of Excellence for Aviation Operations Research (NEXTOR), under FAA cooperative agreement number 01CUMD1. Opinions expressed herein do not necessarily reflect those of the FAA or the U.S. Department of Transportation.

REFERENCES

- [1] M. Adams, S. Kolitz, J. Milner, and A. Odoni. Evolutionary concept for decentralized air traffic management. *Air Traffic Control Quarterly*, 4:281-306, 1997.
- [2] Gaertner, N., M. Klopfenstein and G. Wilmouth, Updated Operational Concept for System Enhancements for Versatile Electronic Negotiation (SEVEN), September 7, 2007 available at <http://cdm.fly.faa.gov/index.html>
- [3] Metron Aviation. www.metronaviation.com/airspace-programs.php
- [4] H. Moulin. The proportional random allocation of indivisible units. *Social Choice and Welfare*, 19:381-413, 2002.
- [5] H. Moulin and R. Stong. Fair queueing and other probabilistic allocation methods. *Mathematics of Operations Research*, 27:1-30, 2002.
- [6] H. Moulin and R. Stong. Filling a multicolor urn: an axiomatic analysis. *Game and Economic Behavior*, 45:242-269, 2003.
- [7] A. Odoni and T.P. Fan. The potential of demand management as a short-term means of relieving airport congestion. In *ATM-2001 EUROCONTROL-FAA Seminar, Santa Fe, NM*. 2001.
- [8] B. O'Neill. A problem of rights arbitration from the talmud. *Mathematical Social Sciences*, 2:345-371, 1982.
- [9] S. Rassenti, V. Smith, and R. Buffin. A combinatorial auction mechanism for airport time slot allocation. *Bell Journal of Economics*, 13:402-417, 1982.
- [10] N. Vakili Pourtaklo, *Preference-Based Fair Allocation of Limited Airspace Resources*, PhD - Electrical Engineering, U of Maryland, December, 2010.
- [11] N. Vakili Pourtaklo, M. Ball. Equitable allocation of enroute airspace resources. *Eighth USA/Europe Air Traffic Management Research and Development Seminar*, 2009.
- [12] T.W.M. Vossen *Fair Allocation Methods in Air Traffic Management*. PhD Thesis, University of Maryland, R.H. Smith School of Business, 2002.
- [13] Vossen, T., M. Ball. 2006. Optimization and mediated bartering models for ground delay programs. *Naval Research Logistics*, 53:75-90.
- [14] M. Wambganss. Collaborative decision making through dynamic information transfer. *Air Traffic Control Quarterly*, 4:107-123, 1996.
- [15] H.P. Young. Disruptive justice in taxation. *Journal of Economic Theory*, 48:321-335, 1988.
- [16] H.P. Young. Progressive taxation and equal sacrifice. *American Economic Review*, 80:253-266, 1990.

AUTHOR BIOGRAPHY

Nasim Vakili Pourtaklo received her PhD in electrical and computer engineering from University of Maryland. She received her Bs. in electrical engineering from Sharif University of Technology. She worked as a project manager at Microsoft. She is a researcher at the NEXTOR lab where she is working on resource allocation and optimization projects.

Michael Ball is the Orkand Corporation Professor of Management Science in the Robert H. Smith School of Business at the University of Maryland. He also holds a joint appointment within the Institute for Systems Research (ISR) in the Clark School of Engineering. Dr. Ball received his PhD in Operations Research in 1977 from Cornell University. He is co-Director of NEXTOR, the National Center of Excellence for Aviation Operations Research, and he leads the NEXTOR Collaborative Decision Making project.

Modeling and Predicting Taxi out Times

Yu Zhang and Arjun Chauhan

Department of Civil and Environmental Engineering
 University of South Florida
 Tampa, Florida, USA
 yuzhang@eng.usf.edu

Xing Chen

Federal Aviation Administration
 Washington DC, USA
 Xing.chen@faa.gov

Abstract — This paper proposes a set of regression equations to model the taxi-out and taxi-in times at airports. The estimated results can be used to calculate the nominal taxi times, which are essential measures for evaluating the taxiing delays at airports. Given the outcomes of the regression model, an iterative algorithm is developed to predict taxi times with inputs such as gate out times, landing times, and runway capacities. A case study at LGA shows that the proposed algorithm demonstrates a higher accuracy in comparison to other algorithms in existing literature.

Keywords-taxi time delay; nominal taxi times; predicting taxi-out time; iterative algorithm.

I. INTRODUCTION

The rise of urbanization has taken its toll on the airline industry among many others. There has been consistent increase in airline traffic from the time it began. Today there are about 7000 flights in America's skies during the peak hours. This is despite a slump in air traffic recently due to the global market meltdown. The air traffic has still been up when compared to the periods before the recession. This rise in airline traffic has seen major delays in the National Airspace System (NAS). A large percentage of flight delay is due to ground holding and ground transit, which includes taxiing delay [1]. Taxi times are the times spent by an aircraft between rolling from a gate to the end of a runway where it takes off or from the entrance of taxiways to a gate after it lands on a runway. Taxiing-in and out are major parts of arrival and departure processes. Considering the distribution of delays experienced by a flight, taxi out delay contributes to 26 percent of the total. According to BTS, 2007 has been a year of the highest taxi times recorded that surpassed the previous peak in the year 2000 [2]. The average block times between busy city pairs in the U.S. increased accordingly, for example, according to Air Transport Association (ATA), in New York LaGuardia (LGA) – Ronald Reagan Washington National (DCA) route segment, the average block time grew by nine minutes from 1995 to 2005 [3]. Longer taxi times have elevated the direct operating and maintenance costs as well as negative environmental impacts in terms of amplified noise and augmented air pollution on and around the airport.

To mitigate delay problems, the FAA implemented the Collaborative Decision Making (CDM) approach in 1998. The CDM in the US is intended towards improving air traffic flow issues in the National Airspace System (NAS) through exchange of information among the air traffic flow managers, air traffic controllers, and airlines. In the US, the initial focus

of the CDM was the Ground Delay Program Enhancements (GDP-E) where the airlines share flight cancellation and reordering information with the Air Traffic Control System Command Center. The users of the NAS also use CDM tools to share information on safety and efficiency among themselves. The CDM concept applied to some EU airports is known as Airport CDM (A-CDM) [4]. The focal point of A-CDM is to bring together the major airport partners like air traffic controllers, aircraft operator, ground handlers and share data in a clear manner. This becomes significant to achieve a common situational understanding consequently leading to better decision-making processes.

Presently, the Next Generation Air Transportation System (NextGen) is under way, the objective is to improve the NAS to meet future demand, avoid congestion, and make the skies safer. NextGen suggests using various technologies, equipment, and procedures to enhance pilots' control over flight paths while the controllers on the ground focus more on traffic flow management [5]. NextGen looks to implement new tools that are being developed to help manage aircraft flow at airports in order to mitigate taxiing delays, reduce engine run times and consequent environmental impact. Such new tools require a better understanding of the taxi times, taxiing delays, and also call for a way to accurately predict taxi times. Accurate prediction of departure taxi times are essential and help airlines manage push back times, obtain and pass on delay information to destination airports. Correct prediction is a key component of the CDM operations and leads to better gate management and reduced arrival and departure delays. The Air Traffic Control (ATC) will benefit as well via improved demand forecasts for airports and enroute air sectors.

This study contains two parts. In the first part, a set of linear models are established to model the taxi-in and taxi-out times. Besides offering inputs for the predicting model in the second part, the set of linear models can be used to calculate the nominal taxi times, which are essential measures that can be used to evaluate the taxiing delays at the airports. In the second part, an iterative algorithm is proposed to predict the taxi-in and taxi-out time with the outcomes from the regression models and other inputs. In comparison to other existing taxi time predicting model, the outcomes of the case study with our model provide higher accuracy and reliability.

II. LITERATURE REVIEW

The existing model for estimating unimpeded taxi times recorded in the Aviation System Performance Matrix (ASPM)

database is developed by Kondo based on two linear equations, one for taxi-in, the other for taxi-out, while containing both taxi-in and taxi-out queue lengths [6]. Given the actual flight information, such as, actual departure and wheel-off times, Kondo sets up bins for each minute of a single day and count how many departing aircraft ahead of one flight at the queue entry time (gate out time). The number of aircraft ahead is considered as the departure queue length for that flight. Arrival queue length can be obtained in a similar way but considering wheel-on and gate in times. For each group, defined according to carrier and season, the taxi-out time is then modeled as the linear combination of an intercept, weighted taxi-out queue length, and weighted taxi-in queue length, as well as the taxi-in time with a different set of coefficients. Given the recorded data, the intercept and weights (coefficients) can be regressed with Ordinary Least Square method. Assuming the interested flight is the only aircraft moving in the taxiway systems, the nominal taxi-out times are calculated with the regression results and by setting the departure queue length as 1 and arriving arrival queue length as 0. Similarly, the nominal taxi-in time is obtained by setting the number of arriving queue length to be 1 and departing queue length to be 0 in the equation. This model captured the major factor contributing to taxi times, the queue lengths of arrival and departure flights. However, it did not consider other factors such as runway configurations, weather impact, and others.

Causal factors identified in Idris et al's [1] paper include runway configuration, airline/terminal location, departure demand, departure queue size, weather, and downstream restrictions. They stated that the runway configuration determines the flow of aircraft at the airport, presents the level of interaction between the flows, and restricts the capacity of arrivals and departures. Idris et al also discussed weather and downstream restrictions in view of the fact that adverse weather greatly reduces the capacity of the airport. They suggested another way of calculating the arrival and departure queue length, accounting for the passing of aircraft, which is shown in Fig. 1.

Fig. 1 shows four aircraft taxing-out from the gate and taking off. The reference aircraft leaves the gate at a time t_1 and takes-off at a time t_2 . The taxi out duration of the reference aircraft is $t_2 - t_1$. There are three aircraft that have a gate out time before t_1 . However, aircraft 1 takes off at a time after t_2 . This aircraft has been passed by the reference aircraft and will not be counted into the queue length of the reference aircraft. In other words, the departure queue of an aircraft is defined as the number of flights that have a takeoff time during its taxi-out and the arrival queue is defined as the number of flights that have a gate in time falling into its taxi-in duration.

Table I shows an illustrated example of the difference in the calculation of queue lengths from the previous two papers. According to the definition by the FAA Aviation Policy and Plans Office (APO) model the departure queue for NWA at 7:10 am is seven, which is the number of aircraft on the airport surface at its gate out time. The departure queue for that flight is five according to Idris et al.'s definition because it has passed the two flights DAL and FLG that had a gate out time of 7:08 am but took off later than the NWA flight.

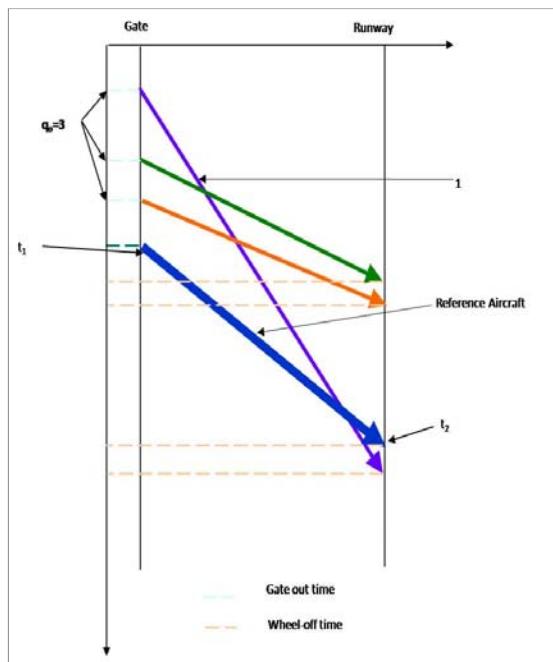


Figure 1. Queue Length Calculation

TABLE I. ILLUSTRATION OF DIFFERENT QUEUE LENGTH CALCULATION

Carrier	Gate-out	Wheels-off	Dep_Queue (Kondo)	Dep_queue (Idris)
USA	6:57:00	7:13:00		
NKS	7:00:00	7:15:00		
NWA	7:00:00	7:18:00		
UAL	7:02:00	7:19:00		
UAL	7:04:00	7:22:00		
DAL	7:08:00	7:29:00		
FLG	7:08:00	7:26:00		
NWA	7:10:00	7:24:00	7	5
AAL	7:14:00	7:27:00		

The queuing model proposed by Idris et al for taxi out estimation assumed takeoff queue to be the primary factor affecting the taxi out time of an aircraft. They set up different combinations of carriers and runway configurations as subsets. The data of the case study that they presented in the paper contained a total of 56 subsets. The downstream restrictions were not considered as separate variables but were assumed to be a part of the departure queue. Idris et al stated that, aircraft that experienced long taxi out times due to passing and restrictions would have long take off queues. For all the subsets, a probability distribution function (PDF) is developed that gives the probability of a queue forming depending on the number of aircraft present on the airport surface at that particular time. An average taxi out time is calculated over all possible queue sizes and then a second-order equation is fitted to these values. Their model was compared to the running average model that is used in the ETMS and showed a reduced mean absolute error. The model predicted 66% of taxi out times within 5 minutes of actual time and is applicable when the number of aircraft present on the airport surface is known.

The Enhanced Traffic Management System (ETMS) model [7] estimates the taxi-out time using the running averages of the last two weeks. The limitation of this model is it does not take into consideration the important factors affecting the taxi-out time of an aircraft such as runway configuration. Shumsky [8] proposed two linear models to predict the taxi-out time of an aircraft. One was a static model and the other was a dynamic one. The static model uses the variables such as carrier, runway configuration, weather and a measure of airport congestion. To explain airport congestion Shumsky projected two different measurements, the number of pushbacks in a given time period around the pushback of the aircraft, and the number of departing aircrafts present on the runway at the pushback time. The results of this study showed that estimations using the queue size were better than using the number of aircrafts on the runway as a measurement for airport congestion. Shumsky also claimed that the static model was as good as the dynamic model for short time horizon, such as, 15-minute period. Nevertheless, for longer time horizon the static model yields superior results.

III. PROPOSED REGRESSION MODEL

This study proposes a set of linear equations to model the taxi-in and taxi-out times. Explanatory variables include arrival and departure queue lengths, runway configuration, arrival and departure runways, and dummy variables indicating time of day and Expect Departure Clearance Time (EDCT) that reflects air traffic flow management activities. Arrival and departure queue lengths and runway configuration have been discussed extensively in the literature review and were widely accepted as major causal factors of taxi-in and taxi-out delay. The information of arrival and departure runways in use are also important because it gives the distance from gates to the end of the runway and the distance from runway exits to gates. Peak and non-peak hours in the day could cause contrasting performance of taxi-in and taxi-out delay due to different gate constraints. In addition, flights experiencing EDCT could perform different from others. Dummy variables are set up for the time of the day and EDCT to account for these effects. Considering the physical interaction between aircraft in the taxiway systems, quadratic terms of the queue lengths are introduced in this regression model. Similar as the APO model, flights are grouped according to carriers and seasons and the flights with taxi times in the upper 25 percent are filtered from the data set as outliers. The case study of this model with 2007 data at LGA shows a higher R square value when compared to other existing models.

A. Explanation of Variables

- *Departure and arrival queues:* These are calculated following the method proposed in Idris et al's paper, which has been described in detail in the literature review.
- *Expected departure clearance times (EDCT):* The traffic management personnel assesses the imbalance of air traffic demand and the capacity of one airport and come up with a plan of holding flights at their origin airports by assigning them expected departure clearance times. Once the EDCT time is allotted the flights have around 15 minutes to depart, otherwise,

they will be assigned a new EDCT time which means more schedule delays. Dummy variable is set to be 1 if one flight experienced EDCT or 0 if it did not.

- *Time of the day:* Peak and non-peak hours have different gate constraints, which per se affect taxi times. After scrutinizing the scheduling, we divide a day at a specific airport into various time windows. For instance, at LGA, we define the different time of the day into four windows, from 6:00am to 9:00am, from 9:00am to 2:00pm, from 2:00pm to 9:00pm, and after 9:00pm. For each time window, dummy variable is set to be 1 if one flight falls into that window or 0 if it did not
- *Runway configuration:* For each runway configuration, the dummy variable is set to be 1 if the configuration was operated while one flight taxiing-in or taxiing-out or 0 if it was not.
- *Arrival and departure runways in use:* Arrival and departure runways in use define the distances from gates to the end of runway(s) and the distances from runway exist(s) to the gates. Nevertheless, this information is hard to obtain. In ASPM data that we used to conduct the case study, there are no arrival and departure runways in use recorded. Fortunately, we can find some airports, LGA as one of them, which has only one arrival runway and only one departure runway. Thus, given the runway configuration, it is easy to know the arrival and departure runways in use. For modeling other airports with more complex runway configuration, additional database, such as Performance Data Analysis and Reporting System (PDARS), need to be used for obtaining such information.

B. Case Study and Data Sources

Airports with longest taxi-out times are typically those with higher volume of air traffic. These airports are mostly either hub airports or focus cities for airlines. According to BTS, for 2007, the top three in the list of airports with longest ground times waiting for takeoff in 2007 were from the New York area and LGA was ranked at number three with average taxi-out times of 29 minutes. As we have described in the Section III-A, not only the runway configuration but also the information of specific runways that flights are assigned to will affect the taxi times. Among the three New York airports, LGA is an ideal airport for our case study because it has only two cross runways, one for arrival and one for departure. The data for the case study was downloaded from aviation system performance metric (ASPM complete).

C. Regression Results and Comparison

With the same data, we conducted the regressions of our proposed model and the existing model used to calculate the nominal taxi times recorded in ASPM database. The comparison of the performance of the two models is shown in table II. The proposed model has an average R^2 value of 0.758 for taxi-out estimation across all groups while the average R^2

value of the other model is 0.429. In addition, the standard error of the R^2 values for the proposed model is smaller.

TABLE II. A STATISTICAL COMPARISON BETWEEN THE ALTERNATE AND THE EXISTING MODELS

R-square Statistics	Alternate model	Existing model
Mean	0.758	0.429
Standard Error	0.004	0.008
Median	0.753	0.434
Mode	0.738	0.455
Standard Deviation	0.044	0.084
Sample Variance	0.002	0.007
Kurtosis	0.814	-0.120
Skewness	0.627	-0.303

IV. PREDICTING TAXI TIMES

A. Iterative Algorithm

Given the regression results and other inputs from flight scheduling, we propose an iterative algorithm to predict the taxi out times. The basic idea is to revise arrival and departure queue lengths and update the taxi-out times of the flights in each iteration until the difference between two iterations becomes less than the convergence parameter set up at the beginning, i.e. $\frac{\sum_{f_a} (t_i^{(n+1)} - t_i^{(n)})}{F_a} < \varepsilon$ for arrival flights and,

$$\frac{\sum_{f_d} (t_o^{(n+1)} - t_o^{(n)})}{F_d} < \varepsilon \text{ for departures flights.}$$

The pseudo code of the algorithm is as below in fig. 2. Initially the arrival and departure queue lengths are set as zero. The iteration count variable n is set as one and convergence parameter is defined as 0.005. Given the estimated coefficients and other input variables, the taxi-in time and taxi-out times can be calculated. Given gate out times and arrival times, we can calculate departure times for departure flights and gate in times for arrival flights. Assuming there are no gate constraints holding arrival flights from getting a gate, we only check the extra taxi-out times that could cause by departure capacity. The 15-minute airport departure rate (ADR) is used as departure capacity of the airport. With the previous calculation, we can check if the 15-minute ADR is exceeded or not. If exceeded, affected flights are postponed to next 15-minute time window. The same procedure is repeated until no demand exceeds supply in all 15-minute time windows in the day. Assuming there is no over passing, we can calculate the arrival or departure queue lengths and then the taxi-in or taxi-out time for each flight. Compare the two sets of taxi-in and taxi-out times mentioned so far, if the differences are smaller than the convergence parameter, the iterative algorithm stops, otherwise, the iteration counts increase one unit and the iteration continues from calculating the departure times for departure flights and gate in times for arrival flights.

1. Initialization queue length: $x_o^{(0)} \leftarrow 0$ and $x_i^{(0)} \leftarrow 0$, iteration count $n \leftarrow 1$, convergence parameter $\varepsilon = 0.005$
2. Given estimated coefficients from regression model, calculate $t_o^{(n)}$ and $t_i^{(n)}$
3. Given gate out time g_o and arrival time a_i , calculate departure time $d^{(n)} = g_o + t_o^{(n)}$ and gate in time $g_i^{(n)} = a_i + t_i^{(n)}$
4. Check 15-minute total departures. If the capacity (ADR) is exceeded, affected flights are moved to the next time window.
Stops when all ADR constraints are satisfied.
5. Calculate departure and arrival queue lengths $x_o^{(n)}$ and $x_i^{(n)}$, assuming no overpassing
6. Given estimated coefficients from regression model, calculate $t_o^{(n+1)}$ and $t_i^{(n+1)}$
7. Convergence test: If $\frac{\sum_{f_a} (t_i^{(n+1)} - t_i^{(n)})}{F_a} < \varepsilon$ and $\frac{\sum_{f_d} (t_o^{(n+1)} - t_o^{(n)})}{F_d} < \varepsilon$, stop, else $n \leftarrow n + 1$ and go to step 3

Figure 2. Pseudo Code of the Alternative Algorithm

B. Case Study and Performance of the Algorithm

We picked one day in 2007, July 13th, at LGA to test the performance of the algorithm. More experiments should be conducted later to get a more general idea about the performance. It shows that the model is able to predict 74% of taxi-out times within five minutes of the actual times. With a different date set, the model proposed by Idris et al predicted 66% of taxi-out times within five minutes of actual times. Table III lists the descriptive statistics when comparing the predicted taxi-out times (CALTO) and actual taxi-out (ACTTO) times recorded in ASPM data. Fig. 3 demonstrates the comparison of average taxi-out times for different hours of the day. It is observed that in the evening, there are larger discrepancies between predicted taxi-out times and actual taxi-out times. It could be caused by the gate constraints that we have ignored in our iterative algorithm or other factors. To predict taxi times more accurately, it is worth of more investigation by looking into surface movement data, observing the real-time operations at airports, and evaluating the impact of gate constraints on arrival queues.

TABLE III. A COMPARISON BETWEEN PREDICTED AND ACTUAL TAXI-OUT TIMES

Statistics	ACTTO	CALTO
Mean	18.55	18.95
Standard Error	0.23	0.22
Median	18.00	18.38
Mode	12.00	19.32
Standard Deviation	5.48	5.25
Sample Variance	29.98	27.56
Kurtosis	0.00	0.72
Skewness	0.63	0.47

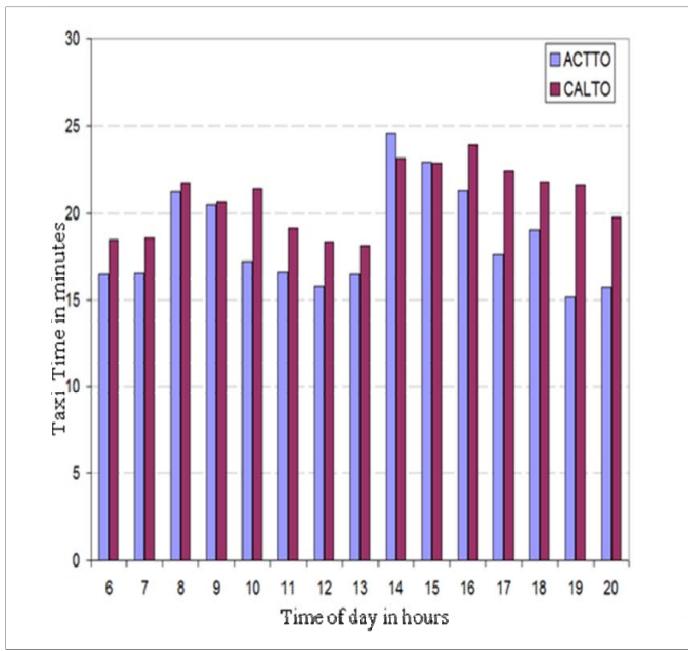


Figure 3. A comparison between Actual and Calculated Taxi times during different hours of the day.

V. CONCLUSION

This paper proposed a set of regression equations to model the taxi times at airports by considering the queuing effect, runway configuration and runways in use, EDCT effect, time of day and others. The comparison of the proposed model and the model used to calculate the nominal times recorded in ASPM database show that with the expansion of independent variables, the proposed model explains double of the variation of the taxi times. The paper then presented an iterative algorithm for predicting taxi times. The inputs for the

algorithm include the estimated coefficients from aforementioned regression model, flight gate out times or arrival times. ADR is taken as the airport departure capacity. Procedures are taken to ensure the departure capacity is not exceeded in each iteration. The algorithm is tested with the data of one day's operations in 2007 at LGA. The predicted results are compared with the actual taxi out times recorded in ASPM. Overall, 74% of predicted value falls into the range within five minutes of the actual times. This is higher than the 66% claimed by one of the existing model, although with data from a different airport.

REFERENCES

- [1] H Idris, J.P. Clarke, R. Bhuva and L. Kang. "Queuing Model for Taxi-Out Time Estimation". Air Traffic Control Quarterly, September 11th 2001
- [2] B. Goldberg and D. Chesser. "Sitting on the Runway: Current Aircraft Taxi Times Now Exceed Pre-9/11 Experience". Bureau of Transportation Statistics Special Report, May 2008.
- [3] J.P. Heimlich. "Commercial Jet Fuel Supply: Impact on U.S. Airlines". Aviation Subcommittee of the committee on Transportation and Infrastructure of the House of Representatives, February 2006.
- [4] Website of the European Organization for the Safety of Air Navigation. <http://www.eurocontrol.int>
- [5] Website of FAA program and initiatives. <http://www.faa.gov/about/initiatives/nextgen>
- [6] A. Kondo. "Derivation of Unimpeded Taxi Times". Office of Aviation Policy and Plans. Federal Aviation Administration.
- [7] ETMS, Enhanced Traffic Management System, Functional Description, Report No. VNTSC-DTS56_TMS-002, Volpe National Transportation Systems Center, U.S. Department of Transportation, Cambridge, MA, 2000.
- [8] R.A. Shumsky. "Dynamic and Statistical Models for the Prediction of Aircraft Take-Off Times". Ph.D. Thesis Operations Research Center, MIT, Cambridge, MA. 1995.
- [9] M.O. Ball, R.L. Hoffman, D. Knorr, J. Wetherly and M. Wambsganss. "Assessing the benefits of Collaborative Decision Making in Air Traffic Management". Air Management R&D Seminar, June 2000.

A TMA 4DT CD/CR causal model based in path shortening/path stretching techniques

Zúñiga C.A.,* Piera M.A.,** Ruiz S.***

Faculty of Telecommunication and Systems Engineering,
Autonomous University of Barcelona
Barcelona, Spain

*catyaatziry.zuniga@uab.cat, **miquelangel.piera@uab.es,
***sergio.ruiz@uab.cat

Del Pozo I.****

Advanced Trajectory Technologies
Boeing Research & Technology Europe
Madrid, Spain

****Isabel.delPozodePoza@boeing.com

Abstract— In the present paper a discrete event model for Conflict Detection and Conflict Resolution algorithm in a TMA 4D trajectory scenario is presented which focuses mainly on the arrival phase. This arises from the overcrowding of airspace near large airports and the need to more efficiently land and take off larger numbers of aircraft. Some attempts to alleviate airspace congestion such as the reduced vertical separation minima, negotiation of voluntary reductions in scheduled service, and the construction of additional runways at major airports, have been done, however, there is still a pending matter to be solved regarding how to improve available airspace capacity avoiding non efficient procedures such as the use of holding trajectories. A deep knowledge about all the events that take place in the management of 4DT and their interactions in a TMA is essential to remove non-effective operations, to avoid delay propagation between arrivals and optimize the occupancy of the runway. The causal model developed considers different alternative pre-defined turning points for each flight evaluating path shortening/path stretching of all trajectories upwards the merging point in a TMA.

Keywords-component; ATM, trajectories, DSS, CPNs, Conflict Detection, Conflict Resolution.

I. INTRODUCTION

Concerns for airspace exert a growing influence on ATM, especially around airports where airspace congestion is becoming a serious problem at many major airports and will become a more severe constraint, especially at the international hub airports serving major European cities and tourist destinations where their ATM-related operations have not yet been fully integrated into the overall ATM organization [5,7].

In the approach phase at the conventional operating methods the Air Traffic Control (ATC) further vector the aircraft to fine tune the sequence and integrate traffic flows from different Initial Approach Fixes (IAFs) to the runway axis avoiding unnecessary gaps at the runway threshold. Flexibility is without doubt one of the main characteristics of this method but unfortunately as a consequence of the strategy followed by controllers for managing arrivals in approach (with the objective of giving themselves more time and margins to make the implementation and fine tuning of the sequence easier) in

high traffic load conditions, often results in high workload both for flight crews and controllers. As a consequence it is evident a difficulty to optimise vertical profiles and to contain the dispersion of trajectories. Numerous actions are required to deviate aircraft from their most direct route for path stretching and later put them back towards a waypoint (e.g. IAF) or the runway axis for integration. [2,8].

With the use of the Flight Management System (FMS)/autopilot-coupled, aircraft are able to fly Required Navigation Performance (RNP) achieving, accurately their desired horizontal paths and flying efficient Continuous Descent Approaches (CDA) from cruise altitudes to touchdown. ATC constraints and ad hoc vectoring in the airspace surrounding the airports limit the ability of these aircraft to effectively use their onboard avionics due to a lack of appropriate traffic planning [9].

An efficient landing sequence will contribute to maintaining the throughput as close as possible to the available runway capacity, ensuring optimisation of the airport manoeuvring area traffic flows and the minimisation of ground and airborne delay while conforming to the separation requirements and will also enable the more widespread use of CDAs [6]. Without an automation planning and decision support tool, it is difficult to accurately predict arrival schedules that are essential for realizing end to-end benefits of RNP for the users flying optimum paths, and for the service providers to manage traffic with varying capabilities with minimum air/ground communications [9].

An innovative technique to tackle the airspace congestion has been developed by the EUROCONTROL Experimental Centre called Point Merge [2]. The Point Merge (PM) technique aims at optimising the use of available airspace in terms of capacity, environmental aspects, and where possible in terms of track distance flown [6]. Point Merge is a structured technique for merging arrival flows derived from an earlier study on airborne spacing sequencing and merging. It is based on a specific route structure (denoted Point Merge System) that is made of a point (the merge point) with pre-defined legs (the sequencing legs) equidistant from this point for path stretching/shortening [10].

In response to the need for an alternative model to build up an efficient landing sequence and conflict free, simulation models could help to analyze the operational efficiency of the current ATC procedure together with airport operations and propose new procedures to optimize the use of available airspace in terms of capacity, environmental aspects, and where possible in terms of track by a proper integration of flows in busy traffic periods [4].

The ATLANTIDA project is a research effort in the air traffic management domain leaded by BR&TE together with key research companies and spanish universities, with the aim of providing a significant contribution to the attainment of the common European goals set by the SESAR program and beyond. Universitat Autònoma de Barcelona (UAB) has been collaborating with Boeing Research & Technology Europe, ATOS-Origin and INDRA in the development of a causal model to improve CD/CR algorithms performance.

The conflict resolution consists in avoidance maneuvers applied by the concerned aircraft. These maneuvers can be heading angle changes (i.e. horizontal deviation), velocity changes, or vertical maneuvers, such as flight level changes for stable aircraft. In a landing sequence, each aircraft concerned computes a conflict-free trajectory for itself. An aircraft is considered in conflict if there exists an instant when the vertical separation and the horizontal separation minima is lost [12].

A simulation model that could cope with the TMA airspace capacity should integrate and manage different sources of information to analyze the perturbations that affect the different arrival flows and design mitigation mechanism to avoid the propagation of those perturbations on the runway throughput. Thus, the model should consider data such as:

- Number of aircrafts in the TMA arrival flows.
- Expected trajectories profile: waypoint pass time, speed, weight.
- Maintaining expected minimum separation standard (MSS);
- Current state of aircraft (level flight; altitude; speed; passing time);
- Geometry information: merging points, entry points, CDA profile for the different aircrafts.

This paper proposes a discrete event model in Coloured Petri Nets (CPNs) for merging arrival flows in an optimal and conflict free sequence of landing aircraft that deals with similar ideas as the Point Merge (Eurocontrol). One of the most important similarities between both approaches is the use of a pre-defined route structure: Terminal Control Area (or Terminal Manoeuvring Area) (TMA) trajectories are characterized by one IAF and a sequence of waypoints some of which are used as merging points with other arrival trajectories. In order to preserve safety distances at merging points speed adjustments must be properly evaluated which are implemented by means of a path stretching/shortening technique. Thus, CPN model compute the exact turning points, (see Fig. 1) according to the type of aircraft (heavy/medium/light), the entry time at the IAF, the expected

speed and the safety distance that should be preserved at the merging point.

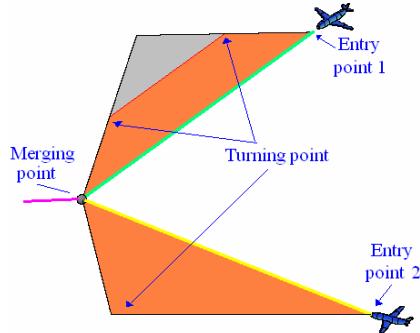


Figure 1. Example of new trajectory with turning point.

In section II the problem scenario is presented and an overview of the conflict detection and resolution algorithm proposed is explained. Section III summarizes the main aspects behind Coloured Petri Nets (CPNs) algorithm. Section IV provides the model description proposed and finally a summary and conclusions are presented in section V.

II. PROBLEM SCENARIO

In the proposed Air Traffic Management System architecture, each aircraft follows a nominal path from source airport to destination airport described by a sequence of waypoints which are fixed points in the airspace. Furthermore, each fixed point has attached a time stamp which represents the expected aircraft pass time through the waypoint (Four dimension trajectories, 4DT's). This pre-defined route structure will be call route. The proposed algorithms are based in a given a sequence of aircraft entering by three different IAFs, flying by its corresponding three routes and a single landing runway.. This routes must join into one route as shown in Fig. 2, by merging in two different merging points.

In the present model, the Gran Canaria TMA is used to test the CPN model (see Fig. 2). There are three different approaching routes from Europe (IAF 1-Rwy, IAF 2-Rwy and IAF 3-Rwy); two different intersection waypoints called *Merging point 1*(*Fayta*) and *merging point 2*(*Cannis*). If the approaching is done by IAF 2 (*Rusik*) or 3 (*Nwpt*) then the *Merging point 1* is the first waypoint where these two routes fuse into one. If the approach is done by IAF 1(*Terto*) then the merging point is *Merging point 2*; in here the three routes (two previously fused), fused again (see Fig. 2).

Conflict detection (CD) is addressed in the proposed model in the following way: The time stamp of the leading and the following aircraft are compared when passing into a merging point to verify if safety distance (sd) is preserved.

As soon as the aircraft arrives at its corresponding IAF, a certain control action is evaluated to guarantee that the aircraft will arrive at the merging point, exactly at a safety distance (in time)of its heading aircraft. The safety distance in the merging point is evaluated independently of the route of each aircraft but taking into account the characteristics of each aircraft (weight). It is important to note that the new expected arrival time is computed at the arrival of the aircraft at the TMA, thus, control actions can be taken at the TMA entry point for each

particular aircraft according to the delay to be generated or absorbed.

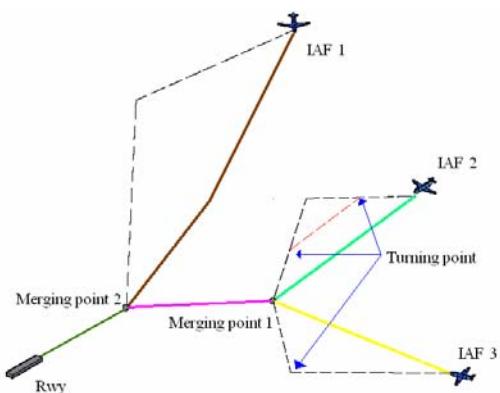


Figure 2. Three trajectories with two merging points.

The predictive model computes and detect Medium Term Conflicts each time there is a new TMA arrival at any entry point, just by evaluating the passing time at the merging points and checking if the time between two consecutive aircrafts is smaller than the minimum separation standard (MSS) given in Table I.

TABLE I. MINIMUM SEPARATION STANDARDS (MSS). MINIMUM TIME BETWEEN SUCCESSIVE ARRIVALS AND DEPARTURES

Leading aircraft	Following aircraft					
	Heavy		Medium		Light	
Heavy	96	60	120	90	144	120
Medium	72	60	72	60	96	90
Light	72	60	72	60	72	60

Control actions that are implemented into the DES model to avoid a conflict (called conflict resolution: CR) are divided in three general aspects.

- A. Evaluate an alternative trajectory (also called change of vector).
- B. To speed up the aircraft.
- C. To decrease the speed of the aircraft.

All control actions are performed by the following aircraft in a look ahead perspective and all are explained bellow. Fig. 3 summarizes the different control mitigation actions implemented as events that can be fired according to time stamp information in the merging points.

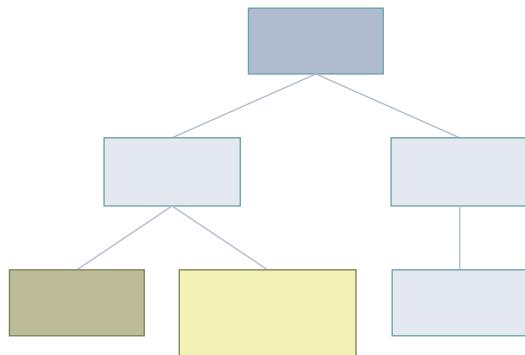


Figure 3. Control actions.

A. Path shortening/Path stretching control action

The objective of the change of vector is to stretch distance to be flown from the TMA to the merging point so an aircraft can arrive at the expected time maintaining its speed profile. Therefore, the delay required is absorbed by the alternative trajectory proposed.

In these algorithms the first approach to solve conflicts starts by modifying the trajectory to be flown. If the difference between the passing time of the leading and the following aircraft is less than the MSS (a conflict is detected) then the following aircraft change its original to modify the passing time to a later time so conflict is avoided.

The delay (in time) needed to arrive on time at the merging point is calculated by comparing desired arrival time at the merging point minus the expected arrival time at the same merging point. This delay obtained is used to calculate the new distance to be flown (see Fig. 4).

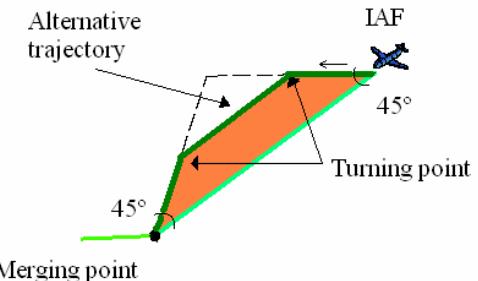


Figure 4. Alternatives trajectories.

To generate the alternative trajectory the new distance to be flown is applied straightforward to a predefined conflict resolution route as shown in Fig. 4. Thus, according to the delay that should be applied to satisfy the desired arrival time at the merging point, and preserving at the same time the speed profile, the aircraft turns 45° until a certain turning point in which the aircraft is redirected (could be parallel to the original route) towards the merging point. The geometry of the alternative trajectory can be triangle or trapezium shaped (see Fig.4).

B. Speed up control action

The objective of the speed up of the following aircraft is to reduce time to be flown from the TMA entry point to the merging point so and aircraft can avoid runway idleness or

future conflicts but ensure a safety distance based on the MSS. By considering that distances between IAFs and the merging points are known variables, the nominal speed profile can be computed straightforward just using the desired time arrival at the merging point.

The following aircraft is speed up until the MSS (according to Table I) is reached or as close to this one as possible (called best separations distance). Then, the best separation distance is the minimum possible distance between the leading and the following aircraft that depends on medium speed (not exceeding a maximum or minimum speed of each aircraft) and the initial time in the TMA that guarantees there is no conflict between them.

C. Decrease speed control action

The objective of decreasing speed of the following aircraft is to augment time to be flown from the TMA to the merging point. This alternative will be used only when the delay obtained through the stretching technique can not solve the conflict detected at the merging point, and an extra delay is required.

III. COLOURED PETRI NET OPTIMIZATION APPROACH

Coloured Petri Nets (CPN) have proved to be successful tools for modelling complex systems due to several advantages such as the conciseness of embodying both the static structure and the dynamics, the availability of the mathematical analysis techniques, and its graphical nature.

The main CPN components that fulfill the modelling requirements are:

- Places: They are very useful to specify both queues and logical conditions. Graphically represented by circles.
- Transitions: They represent the events of the system. Graphically represented by rectangles.
- Input Arc Expressions and Guards: Are used to indicate which type of tokens can be used to fire a transition.
- Output Arc Expressions: Are used to indicate the system state change that appears as a result of firing a transition.
- Colour Sets: Determines the types, operations and functions that can be used by the elements of the CPN model. Token colours can be seen as entity attributes of commercial simulation software packages
- State Vector: The smallest information needed to predict the events that can appear. The state vector represents the number of tokens in each place, and the colours of each token.

The Colour sets will allow the modeller to specify the entity attributes. The output arc expressions will allow specifying which actions should be coded in the event routines associated with each event (transition). The input arc expressions will allow specifying the event pre-conditions. The state vector will allow the modeller to understand why an event can appears, and consequently to introduce new pre-conditions (or remove

them) in the model, or change some variable or attribute values in the event routines to disable active events.

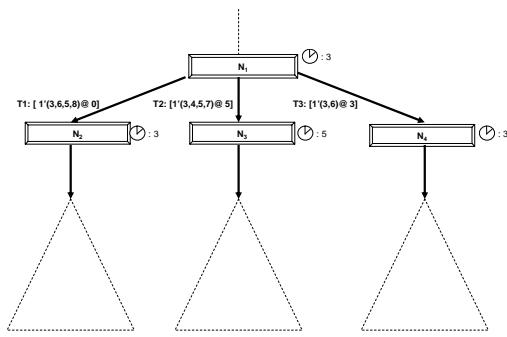


Figure 5. First 2 levels of a coverability tree.

From the OR point of view, the CPN model can provide with the following mathematical structures:

- Variables: A variable can be identified for each colour specified in every place node.
- Domains: The domains of the variables can be easily determined by enumerating all the tokens specified in the initial state.
- Constraints: Can be obtained by straightforward from the arc and guard expressions. Arc expressions can contain constant values, colour variables or mathematical expressions.

From the AI point of view, the coverability tree of a CPN model allows to determine:

- All the events that could appear according to a particular system state (Fig. 5).
- All the events that can set off the firing of a particular event.
- All the system states (markings) that can be reached starting from a certain initial system operating conditions M_0 .
- The transition sequence to be fired to drive the system from a certain initial state to a desired end-state.

IV. MODEL DESCRIPTION

The medium term CD&CR model proposed has been specified in the Coloured Petri Net (CPN) formalism. The discrete event approach has been specified using seven colours, five places and nine transitions. This model can be integrated with the AMAN/DMAN CPN model [6] to optimize a shared mode runway in which the best landing sequence can also be computed.

As shown in Fig. 3, three different control actions attached to each merging point could be fired. These events are represented by different transitions in the CPN model.

- The event “Change trajectory” takes place when a conflict between two aircraft is detected in a merging point. Taking into account the scenario presented in Fig. 2, three transitions derivates from this event:

Change trajectory from IAF 1 to merging point 2, change trajectory from IAF 2 to merging point 2 and finally change trajectory from IAF 3 to merging point 2.

- The event “Change trajectory + decrease speed” takes place when a conflict between two aircraft is detected in a merging point and can not be solved only by a changing vector procedure. Therefore, taking into account the scenario presented in Fig.2, three transitions derivates from this event: reduce speed from IAF 1 to merging point 2, reduce speed from IAF 2 to merging point 2 and finally reduce speed from IAF 3 to merging point 2.
- The event “Speed up” takes place when the separation distance between the leading and the following aircraft is greater than the MSS. In order to acquire a reduction of time from the TMA entry point to the merging point the following aircraft is accelerated until an optimal or best separations distance (in time) is reached. Another transition is used to compute the speed profile along the new trajectory so no extra speed changes will be required.

A. Net specification & description

Table II summarizes the colours used to describe all the information required in the places to define the aircraft trajectory in 4D.

Place specifications are shown in Table III and detailed as follows: In the CPN representation, “Segments” place node has information regarding each aircraft trajectory sucha as: Colour *aid* corresponds to the aircraft identification in a trajectory; the first *idp* colour corresponds to the waypoint identification of the beginning of the trajectory (entry point in the TMA) while the second *idp* keeps the information regarding to the passing waypoint identification; colour *t* and *vel* carry on its corresponding current time and speed. The third *idp* colour corresponds to the next waypoint identification of the trajectory while second colour *t* has information about IAFs entry time. Finally the *t* and *de* colour corresponds to IAFs entry time and distance in the TMA, respectively.

TABLE II. COLOUR SPECIFICATION

Colour	Meaning
<i>aid</i>	Aircraft identification
<i>idp</i>	Waypoint identification
<i>t</i>	Time
<i>vel</i>	Average speed
<i>de</i>	Distance between two waypoints
<i>c</i>	Control variables
<i>wp</i>	Waypoint information

TABLE III. PLACE SPECIFICATION

Place	Colour	Definition
Segments	S	<i>aid</i> * <i>wp</i> * <i>wp</i> * <i>t</i> * <i>v</i> * <i>wp</i> * <i>t</i> * <i>de</i>
G	G	c,c,c
Solution	R	<i>aid</i> , <i>wp</i> , <i>v</i> , <i>de</i>
Pair	P	<i>aid</i> , <i>aid</i> , <i>t</i> , <i>wp</i> , <i>wp</i>

Place “G” considers only three colours as shown in Table III. The first colour (ie. *d10*) *takes a value 0* only when the pair of consecutive aircraft has not previously evaluated or they have been forced to change the speed in order to avoid conflicts. The same colour takes value 2 when the pair of aircrafts has to be evaluated in merging point 1 after a vector change (path stretching) has been proposed. Finally, colour *d10* *takes value 3* when the pair of aircrafts has to be evaluated in merging point 2 after a speed-up control action has been applied. Colour *c3* and *c4* are control variables that indicate the number of the following aircraft to be evaluated in Fayta and Canis, respectively. Therefore if the transition concerns to Fayta, colour *c3* will be incremented in one unit to update the next pair to be evaluated in this waypoint.

Place “Solutions” store the information regarding to the aircraft successfully solved. This node contains the aircraft identification, its corresponding passing time, medium speed and distance to be flown.

Place “Pair” contains the information that links the leading and the following aircraft (‘*x1*’ and ‘*y1*’, respectively) with the next passing time (*x11*) of aircraft ‘*y*’ and a control variable (*c3* or *c4*) that indicates the following aircraft to be evaluated in Fayta or Canis. After firing this transition the information is properly updated in all place nodes since the vector change has been completed.

V. EVENT SPECIFICATION EXAMPLE

Fig. 6 illustrates an event (represents an aircraft that will be speed up) that formalize the CD&CR model for merging point 1.

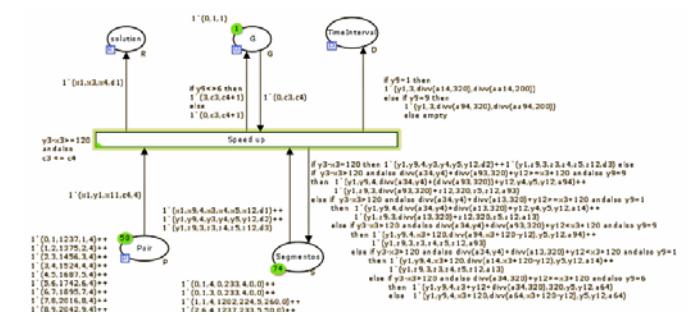


Figure 6. Example of the CPN for the CD&CR algorithm.

The CPN shows five nodes; the node “Segments” asks as initial conditions for a pair of aircraft (*x1* & *y1*) that comes from their corresponding passing waypoint identification (*x9*, *y9* & *z9*) and they are evaluated when passing in merging point

$1 (x_2=y_2=4)$ and with passing time, speed profile, next waypoint identification, IAFs entry time and distance from IAFs (x_3,x_4,x_5,x_{12},d_1 & y_3,y_4,y_5,y_{12},d_2 & z_3,z_4,z_5,z_{12},d_3 respectively).

As initial conditions, node “*Pair*” asks for the same pair of aircraft that node “*Segments*” (x_1 & y_1), with a next passing time (x_{11}), c_4 as a control variable to indicate the following aircraft, and finally $c_3=4$ indicating they are evaluated when passing in merging point 1).

Information supported in node “*G*” is used when colour d_{10} takes value 0 since the following aircraft has not been previously evaluated, and c_3,c_4 are linked to node “*pair*”.

When all initial conditions are properly specified in each place node, then node “*Solutions*” stores only information regarding the leading aircraft (note that this aircraft will not have conflict with any other aircraft). Node “*G*” will increase colour value c_4 in one unit to specify that the next aircraft has changed and should be evaluated; and if the following aircraft comes from IAF 2 or 3, $d_{10}=3$ in order to be re-evaluated in merging point to solve any possible conflict in this passing waypoint.

Finally place “*Segments*” will return information about the following aircraft with its corresponding new passing time in merging point 1, new speed profile , and/or new distance to be flown, if required.

VI. CASE STUDY

Arriving flow to the Gran Canaria TMA landing at Gran Canaria airport at a busy traffic period use to be between 20 to 30 aircraft in one hour. To test the performance of the proposed CD/CR CPN model, a synthetic traffic workload of 35 arrival aircrafts has been designed. The arrival traffic sequence is assumed to have been determined upstream in the extended TMA by an arrival manager (AMAN). Therefore sequencing is implicitly defined in the traffic preparation input file.

Aircraft arrivals through the same entry point are conflict-free between them; however the merging of these arrivals generate conflicts at the merging point areas.

Table IV illustrates the 4DT specification of the first three aircraft arriving to Gran Canaria TMA. As it can be noted, 2 aircraft arrive through Rusik entry point and one aircraft arrives through Terto entry point. The aircrafts arriving through Rusik are conflict free between them, but there is a conflict in Cannis merging point.

TABLE IV. TWO TRAJECTORIES EXAMPLE

No	TMA entry time (s)	TMA IAF	WPT merging point 1	WPT time (s)	WPT TAS (m/s)	WPT merging point 1	WPT time (s)
1	260	2	3	972	224	4	1202
2	50	1	3	---	233	4	1237
3	470	2	3	1154	233	4	1375

According to this information in Table IV, the initial marking for places has the format shown in Table V.

TABLE V. INITIAL MARKING FOR TWO TRAJECTORIES EXAMPLE

Place	Initial marking
Segments	$1^*(1,1,3,972,224,260,0)+1^*(1,1,4,1202,224,260,0)+1^*(2,6,3,0,233,50,0)+1^*(2,6,4,1237,233,50,0)++1^*(3,1,4,1375,233,5,470,0)+1^*(3,1,3,1154,233,4,470,0)$
Pairs	$1^*(0,1,1237,1,4)+1^*(1,2,1375,2,4)+1^*(2,3,1456,3,4)$
G	$1^*(0,1,1)$
Solution	empty

For these three trajectories a feasible conflict free solution is reached (Table VI). It can be noticed that the aircraft number 1 has been accelerated (to 290m/s) and has new waypoint passing time in merging point 1 of 810s instead of 972s and in merging point 2 the passing time has changed from 1202s to 988s, as a result aircraft 3 has also a new waypoint passing time of 1043 instead of 1154 in merging point 1, and its corresponding speed has also been changed to 278m/s.

TABLE VI. SOLUTION FOR TWO TRAJECTORIES EXAMPLE

No	TMA entry time (s)	TMA IAF	WPT merging point 1	WPT time (s)	WPT TAS (m/s)	WPT merging point 1	WPT time (s)
1	260	2	3	810	290	4	988
2	50	1	3	---	290	4	1108
3	470	2	3	1043	278	4	1228

The entirely model has been tested with 35 aircraft and the results are shown in Table VII (see the Appendix A). Information about aircraft are black colored while the results are in red color to be identified easily.

At the right hand side of figure 7, a trombone area representing the PM approach with two entry points is represented, while at the left hand side of the same figure the fixed re-routes of the new approach are also represented for the same TMA configuration. Thus, with the proposed method, each entry point has associated a fixed re-route which is computed by a turn of 45° from the arrival route at the IAF.

One of the main differences between both methodologies is the geometrical configuration of the model proposed. Figure 1 shows a difference in terms of the use of path shortening or path stretching technique.

Furthermore, the proposed model has been tested using three different IAFs and two merging points, providing excellent results for a traffic peak. In fact the causal CPN model could be extended to different number of IAF and merging point configurations, while multiple point merge systems require the analysis of particular solutions to evaluate the cause effect configuration.

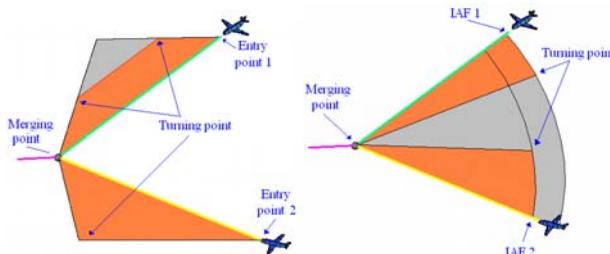


Figure 7. Example of new trajectory with turning point.

VII. CONCLUSIONS

The proposed approach has modelled the CD&CR problem for multi-aircraft using a discrete event approach in the CPN formalism. A case study with 35 arrival aircrafts has been successfully solved.

The CD&CR model computes the future passing times, speed and positions of each aircraft according to certain characteristics (heavy, medium, light). The safety distance requirements due to vortex turbulences at merging points are specified to solve the problem properly. The solution is obtained using the reachability tree in CPN. A computer simulation has used to generate a feasible 4DT solution. The model scope can be extended with consume fuel aspects (BADA referenced) in order to design a cost function that would allow an efficient exploration of the reacheability tree.

One of the key-point of such a design is the use of the state space to understand the behavior of the model. Furthermore, the model has been designed in order to help the modeller to design new procedures to solve problems regarding the future free flight concept.

ACKNOWLEDGMENT

This work is partly funded by ATOS-origin the Science and Innovation Ministry of the Spanish Government "Discrete Event Simulation Platform to improve the flexible coordination of land/air side operations in the Terminal Maneuvering Area (TMA) at a commercial airport" CICYT Spanish program TRA2008-05266/TAIR, and the ATLANTIDA project: "New technologies applied to UAV's for research and ATM development" (CEN20072008)

REFERENCES

- [1] Frazzoli E., Mao Z.H., Oh J.H. Feron Z.E. Resolution of Conflicts Involving Many Aircraft via Semidefinite Programming. *Journal of guidance, control and dynamics*. ISSN 0731-5090, vol. 24, No 1, pp. 79-86, 1999.
- [2] Favennec, B., Symmans, T., Houlihan, D., Vergne, K., and Zeghal, K.. Point Merge Integration of Arrival Flows Enabling Extensive RNAV Application and CDA - Operational Services and Environment Definitio, http://www.eurocontrol.int/eec/public/standard_page/DOC_Report_2008_003.html
- [3] <http://www.eurocontrol.int/prc/index.html>.
- [4] KoSeckA, J., Tomlin, C. Pappas, G. and Sastry, S. Generation of Conflict Resolution Maneuvers for Air Traffic Management. *IEEE Proc. IROS 97* 0-7803-4119-8/ pp. 1598-1603, 1997.
- [5] Kuchar, J.K. and Yang, L.C. A.. Review of Conflict Detection and Resolution Modeling Methods. *IEEE Transactions on Intelligent Transportation Systems*, Vol. (1), No. (4). Pp. 179-189, 2000.
- [6] Piera, M.A., and Baruwa, O.T., A Discrete Event System Model to Optimize Runway Occupancy, INO Workshop, EUROCONTROL 2008.
- [7] Sastry S., Meyer G., Tomlin C., Lygeros J., Godbole D., Pappas G., Hybrid Control in Air Traffic Management Systems. In poceedings of the 1995 IEEE Conference in Decision and Control. pp. 1478—1483, 1995.
- [8] Favennec B., Hoffman E., Trzmiel A., Vergne F., Zeghal K. The Point Merge Arrival Flow Integration Technique: Towards More Complex Environments and Advanced Continuous Descent, In proceeding of 9th AIAA Aviation Technology, Integration, and Operations Conference (ATIO) 21 - 23 September 2009, Hilton Head, South Carolina.
- [9] Mohleji S.C., Stevens R.K., Optimizing flight path for RNP aircraft in busy terminal- first step towards 4D navegation, 25th Digital Avionics Systems Conference, IEEE/AIAA, PP. 1-9 2006
- [10] Ivanescu, D., Shaw, C., Tamvaclis, C., and Kettunen, T., Models of Air Traffic Merging Techniques: Evaluating Performance of Point Merge, In proceeding of 9th AIAA Aviation Technology, Integration, and Operations Conference (ATIO) 21 - 23 September 2009, Hilton Head, South Carolina.
- [11] Ivanescu, D., Shaw, C., Tamvaclis, C., and Kettunen, T., Integrating Aircraft Flows in the Terminal Area with no Radar Vectoring, In proceeding of 9th AIAA Aviation Technology, Integration, and Operations Conference (ATIO) 21 - 23 September 2009, Hilton Head, South Carolina.
- [12] Archambault, N., Durand, N., Scheduling heuristics for on-board sequential air conflict solving, Digital Avionics Systems Conference, DASC 04, The 23rd, Vol. 1, pp. 3.A.1-1 to 3.A.1-9, 2004

VIII. APENDIX A

In this section the case study information can be found. Case study has been with 35 aircraft with initially 10 conflicts detected in the first point merge and 21 more conflicts in the second point merge. The results as well as all the regarding information to the name and nominal passing time of the IAF, nominal speed, distance to be flown, new speed and passing time, are shown in Table VII.

TABLE VII. 35 TRAJEYTOR STUDY CASE

	TMA entry time	TMA IAF	WPT TAS (m/s)	WPT name	WPT TAS (m/s)	WPT name	WPT nominal time (s)	WPT TAS (m/s)	New WPT time (s)	New speed (m/s)	Distances to be flown (m)	WPT name	WPT nominal time (s)	WPT nominal TAS (m/s)	New WPT time (s)	New speed (m/s)	Distances to be flown (m)	WPT nominal time (s)	WPT nominal TAS (m/s)
1	260	RUSIK	224	FTV	224	FAYTA	971,93	224	810	290	159867	CANIS	1202	218,67	988	290	211350	1373,23	191,19
2	50	TERTO	233	LZR	233	BETAN	978,90	233		290		CANIS	1237	221,20	1108	283	300208	1404,94	196,05
3	470	RUSIK	233	FTV	233	FAYTA	1154,47	233	1043	278	159867	CANIS	1375	224,06	1228	278	211350	1541,98	197,55
4	870	NWPT				FAYTA	1224,84	224	1050	274	79651	CANIS	1456	213,84	1348	274	131134	1630,42	188,52
5	180	TERTO	224	LZR	224	BETAN	1250,18	224				CANIS	1524	203,47	1468	233	300208	1706,55	181,41
6	400	TERTO	233	LZR	233	BETAN	1328,90	233				CANIS	1687	221,20	1588	252	300208	1754,94	196,05
7	800	RUSIK	224	FTV	224	FAYTA	1511,93	224	1486	232	159867	CANIS	1742	218,67	1708	232	211350	1913,23	218,67
8	990	RUSIK	233	FTV	233	FAYTA	1674,47	233	1623	252	159867	CANIS	1895	224,06	1828	252	211350	2061,98	197,55
9	1430	NWPT				FAYTA	1784,84	224	1625	253	79651	CANIS	2016	213,84	1948	253	131134	2190,42	188,52
10	1100	RUSIK	224	FTV	224	FAYTA	1811,93	224	1812	224	159867	CANIS	2042	218,67	2068	224	216832	2213,23	218,67
11	1290	RUSIK	233	FTV	233	FAYTA	1974,47	233	1969	235	159867	CANIS	2195	224,06	2188	235	211350	2361,98	197,55
12	1000	TERTO	233	LZR	233	BETAN	1928,90	233				CANIS	2287	221,20	2308	233	304764	2354,94	196,05
13	1800	NWPT				FAYTA	2154,84	224	2155	224	79651	CANIS	2386	213,84	2428	224	140672	2560,42	188,52
14	1180	TERTO	224	LZR	224	BETAN	2250,18	224				CANIS	2524	203,47	2548	224	306432	2706,55	181,41
15	1700	RUSIK	233	FTV	233	FAYTA	2384,47	233	2295	233	138635	CANIS	2605	224,06	2668	233	225544	2771,98	197,55
16	2060	NWPT				FAYTA	2414,84	224	2415	224	82979	CANIS	2646	213,84	2788	224	163072	2820,42	188,52
17	1500	TERTO	233	LZR	233	BETAN	2428,90	233				CANIS	2787	221,20	2908	233	328064	2854,94	196,05
18	1905	RUSIK	224	FTV	224	FAYTA	2616,93	224	2670	224	171459	CANIS	2847	218,67	3028	224	251552	3018,23	218,67
19	1690	TERTO	224	LZR	224	BETAN	2760,18	224				CANIS	3034	203,47	3148	224	326592	3216,55	181,41
20	2200	RUSIK	233	FTV	233	FAYTA	2884,47	233	2924	233	168751	CANIS	3105	224,06	3268	233	248844	3271,98	197,55
21	1900	TERTO	233	LZR	233	BETAN	2828,90	233				CANIS	3187	221,20	3388	233	346704	3254,94	196,05
22	2420	RUSIK	224	FTV	224	FAYTA	3131,93	224	3095	224	151200	CANIS	3362	218,67	3508	224	243712	3533,23	218,67
23	2860	NWPT				FAYTA	3214,84	224	3215	224	91939	CANIS	3446	213,84	3628	224	172032	3620,42	188,52
24	2150	TERTO	224	LZR	224	BETAN	3220,18	224				CANIS	3494	203,47	3748	224	357952	3676,55	181,41
25	2650	RUSIK	224	FTV	224	FAYTA	3361,93	224	3464	224	182336	CANIS	3592	218,67	3868	224	272832	3763,23	218,67
26	2400	TERTO	224	LZR	224	BETAN	3470,18	224				CANIS	3744	203,47	3988	224	355712	3926,55	181,41
27	2900	RUSIK	233	FTV	233	FAYTA	3584,47	233	3685	223	182905	CANIS	3805	224,06	4108	233	281464	3971,98	197,55
28	2600	TERTO	233	LZR	233	BETAN	3528,90	233				CANIS	3887	221,20	4228	233	379324	3954,94	196,05
29	3450	NWPT				FAYTA	3804,84	224	3805	224	121059	CANIS	4036	213,84	4348	224	201152	4210,42	188,52
30	2820	TERTO	224	LZR	224	BETAN	3890,18	224				CANIS	4164	203,47	4468	224	369152	4346,55	181,41
31	3300	RUSIK	224	FTV	224	FAYTA	4011,93	224	3892	224	132608	CANIS	4242	218,67	4588	224	288512	4413,23	218,67
32	3750	NWPT				FAYTA	4104,84	224	4334	212	123915	CANIS	4336	213,84	4708	212	204008	4510,42	188,52
33	3100	TERTO	233	LZR	233	BETAN	4028,90	233				CANIS	4387	221,20	4828	233	402624	4454,94	196,05
34	3600	RUSIK	233	FTV	233	FAYTA	4284,47	233	4604	233	233991	CANIS	4505	224,06	4948	233	314084	4671,98	197,55
35	3300	TERTO	224	LZR	224	BETAN	4370,18	224				CANIS	4644	203,47	5068	224	396032	4826,55	181,41

Impact of Lightning Strikes on National Airspace System (NAS) Outages

A Statistical Approach

Aurélien Vidal

*University of California at Berkeley
NEXTOR
Berkeley, CA, USA
aurelien.vidal@berkeley.edu*

Jasenka Rakas

*University of California at Berkeley
NEXTOR
Berkeley, CA, USA
jrakas@berkeley.edu*

Abstract— Although it is reasonably well accepted that lightning strikes are a significant cause of outages on the National Airspace System (NAS), there remains a serious lack of comprehensive analyses providing sound estimates of outages caused by convective weather. Current estimates and methods generally cover specific outages and try to determine their causes by comprehensively analyzing the lightning strikes that occurred in the vicinity of the system. Such methods are inadequate when trying to evaluate the global impact of convective weather on very large systems such as the NAS, which is composed of more than 70,000 systems. In this paper, a statistical method is developed to estimate the number of outages caused by lightning strikes, which take into account both the time detection of the outage and the localization of the strikes. In addition, we present results of its application on the NAS outages between 1999 and 2005.

Index Terms - component; lightning strikes; outages; NAS; NLDN; logit model

I. INTRODUCTION

The safety and efficiency of air transportation within the United States (US) largely relies on the reliability of the National Airspace System (NAS). Lightning strikes are believed to be one of the major causes of electrical power interruption and occur during critical weather conditions [1] for airborne aircraft.

The method presented here is based on the National Lightning Detection Network (NLDN) observations that represent more than 95% of the lightning strikes in the continental US [2] and provide the time of the strike, its location and the intensity. The database contains the location of only the first stroke of each lightning strike, which can cause errors of up to 10 km, based on the algorithm used by the NLDN [3].

The information concerning NAS outages, provided by the Federal Aviation Administration (FAA), were gathered manually by the technicians in charge of the maintenance of the NAS, utilizing a system of codes to classify the different categories of outages.

A preliminary analysis of the problem was carried out in an attempt to establish an initial relation between lightning strikes and outages. All strikes that happened in the continental US between 1999 and 2005 were attributed to corresponding Air Route Traffic Control Centers (ARTCC) and to a month. Then, values were plotted against the number of outages occurring in the respective ARTCC to determine if a general pattern existed (Figure 1). Although a small positive correlation is noticeable, results are unusable for further applications as in many cases more lightning strikes do not imply more outages. Indeed, more advanced tools based on the characteristics of lightning strikes are required to explain the outage process.

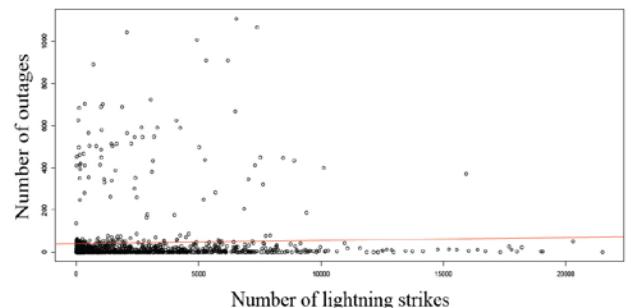


Figure 1. Plot of the number of lightning strikes during a month and inside a given ARTCC against the number of outages inside the same region and over the same period of time

II. CREATION OF THE DATA BASE

A. Data Management

The data used for this study contain all the lightning strikes in the continental US between March 1999 and November 2005 (recorded by the NLDN). The data contain both cloud-to-cloud (CC) and cloud-to-ground (CG) lightning strikes. Beside interferences, the former do not impact the NAS systems significantly and are therefore ignored in the study. For each strike, the data set provides the location, the time and the

magnitude (in kilo amperes) but does not give the number or the location of each stroke.

The list of all the outages that concerned the NAS during the same period constitutes the second data set. For each outage, the information available is: the type of system, its location, the beginning and end of the outage, and reported cause of the failure. The code corresponding to "Convective Weather" is 85-3 (Weather Effect – Lightning Strikes) and all outages with a different code are not used to develop the model. It is interesting to note that the beginning time of the outages is the moment when the outage was detected and not necessarily the time when it occurred. Finally, as the NLDN covers only the continental US, all the outages concerning systems located in the ARTCC of Honolulu (ZHN), Anchorage (ZAN) were also removed.

The number of remaining outages is close to 900.

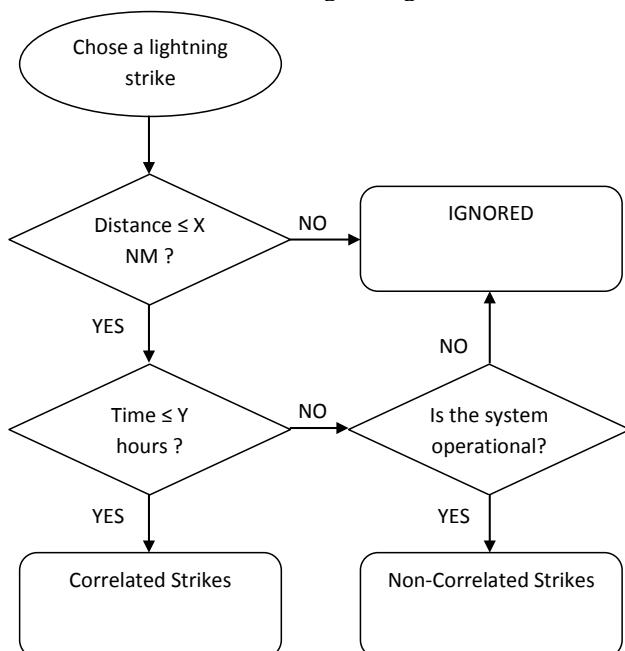


Figure 2. Algorithm applied to each outage and for all lightning strikes

B. Correlated vs Non-Correlated Strikes

The input of the model is built with 2 distinct selections of lightning strikes: one group that caused outages (Target Strikes) and another that occurred in the vicinity of the systems but that did not cause any outages (Non-Correlated Strikes). The former cannot be directly created considering the lack of precision of the NLDN data and therefore a larger (and simpler to define) group of strikes that might have caused an outage is required (Correlated strikes):

- Correlated Strikes: strikes that might be responsible for the outage,
- Target Strikes: strikes that are considered as responsible for the outage,
- Non-Correlated Strikes: strikes that did not cause any outage,
- Undetermined strikes: strikes that do not belong to the three previous categories.

The Undetermined strikes are mainly strikes that occurred in the vicinity of the system when it was inoperable for any reason (maintenance, outage, etc). Strikes are first selected based on a spatial criteria of X nautical miles and then grouped according to a temporal criteria (Figure 2).

C. Time and Space Window

The separation between the different groups of strikes is based on a time criterion and a space criterion. The name used for the study is "Time and Space Window" and corresponds to an X NM radius circle around the system and a period of Y hours before the outage. The final values for X and Y are respectively 5 NM and 48 hours, which resulted from the analysis of the Cause Code 85-3 outages. A good Time and Space Window must give both a high level of correlation (as we expect that most of the Cause Code 85-3 outages are caused by lightning strikes) and a small number of correlated lightning strikes to facilitate the Target Strikes extraction. The level of correlation represents the ratio of the number of correlated events (outages) for a period over the total number of events (outages) for the same period. An outage is considered correlated if its Time and Space Window contains at least one

Table 1: Level of Correlation as a function of the Time Window size and the Space Window size

Distance (NM)	Time									
	5min	30min	1h	6h	12h	24h	48h	72h	96h	120h
0.1	0%	1%	2%	3%	4%	5%	5%	5%	5%	5%
0.25	2%	8%	12%	17%	20%	23%	25%	25%	25%	25%
0.5	6%	16%	22%	32%	37%	42%	46%	46%	47%	47%
0.75	10%	22%	28%	41%	46%	52%	56%	57%	58%	59%
1	13%	24%	31%	44%	50%	57%	62%	64%	64%	65%
2	20%	32%	40%	53%	60%	67%	72%	74%	75%	76%
3	25%	35%	42%	57%	62%	70%	76%	78%	80%	81%
4	28%	38%	44%	58%	65%	72%	79%	82%	83%	84%
5	30%	40%	46%	61%	66%	74%	80%	83%	84%	85%
10	31%	41%	48%	63%	69%	78%	85%	87%	88%	89%
15	32%	42%	49%	64%	70%	79%	87%	89%	90%	91%

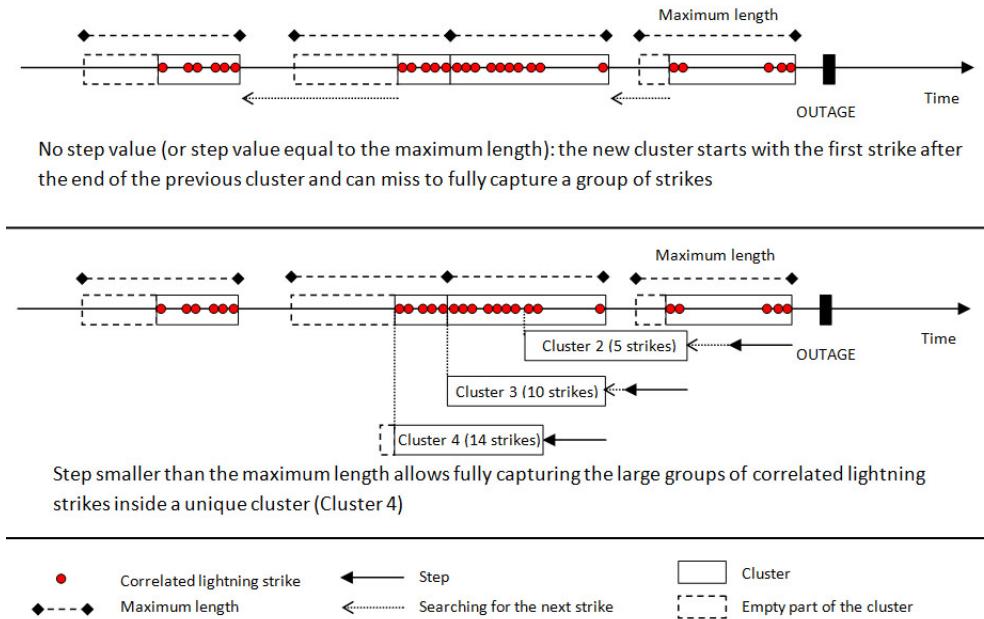


Figure 3: Comparaison between clusters that can overlap and clusters without a step value

lightning strike.

As the Time and Space Window becomes larger, more strikes are correlated. The definition of the window must take into account this tradeoff between the Correlation Ratio and the relevance of the correlated strikes. The Time Window represents the maximum period before an outage during which a strike is considered as a potential cause of outage. Although the outage mechanism is not perfectly understood, it is well accepted that the time between a strike on the system and the failure will be generally very short. Associated failures are typically caused by over intensity resulting in excessive heat inside the system and thus occur almost instantaneously after the impact [3]. As the lightning data and outages data have respectively an average time precision of 5 μ s and 1 minute, a time window of a few minutes should capture most of the cases [3]. The main issue comes from the detection time of the outages. Systems equipped with automatic outage reports are automatically considered as “down” as soon as the outage occurs. For other pieces of equipment such as radars, which are closely monitored, outages are also very likely to be detected within a few minutes. However, for other systems, such as Precision Approach Path Indicator (PAPI), outages can be only detected in the case where someone tries to use the system or during a routine check. Unfortunately, no studies have been carried out to determine the average detection time of outages depending on the type of systems and their location. Different simulations performed with Time windows between 1 minute and 5 days proved that the detection of an outage can take up to several days although more than 50% of the correlated outages have at least one lightning strike in the hour prior to the failure (Table 1).

The Space Window represents the size of the vicinity in which strikes are considered as a potential cause of outage. Its size is the radius of the circle and is centered on the system. In this case, the incertitude mainly comes from the NLDN-stroke-gathering algorithm [3]. Different simulations with space windows between 0.1NM and 15NM showed a similar behavior compared to the Time Window. All curves tend to a limit value after 10 NM. Once again, although some outages have their closest correlated lightning strike located at more than 10 NM, more than 50% of the correlated outages have at least one lightning strike within 0.5 NM (Table 1).

The final size of the Time and Space Window directly determines the correlated strikes, among which the target strikes, input of the statistical model, will be selected. Its definition is based on a tradeoff between the number of observations that should be as large as possible and the relevance of the input. Each time the Time and Space Window increases, new observations are added – with a decreased probability of actually being involved in the outage process. As showed in Table 1, the correlation ratio reaches a limit in both cases. Those values are considered as the maximum number of observations which can be used, with 95% of them being used inside the model. The last 5% of correlation (i) implies a multiplicative factor of 2 to 3 for the time and space parameters, and (ii) adds a large number of correlated strikes with a relatively low relevance.

D. Strike Definition vs Cluster Definition

Strikes are defined as the strikes that caused the outage as opposed to the correlated strikes, which are the strikes that *might have* caused the outage. However, this definition results in many possible solutions as no general rule or model exists to

determine, among a group of strikes, which one actually caused the outage.

Clusters were created to solve the problem of best strikes selection. The idea is to gather correlated strikes over a period of time and then consider the whole cluster as the outage cause. This principle was born from the observation of the time localization pattern of Correlated Strikes that tend to appear in groups of high time density. As lightning strikes happen during thunderstorms, multiple strikes generally hit the vicinity of the system within a short period. The goal of cluster models is to isolate those laps of time of high convective activity and consider them as the cause of the outage.

Clusters are fully defined by 2 different characteristics:

- Maximum Length: their maximum duration in time. Clusters with maximum lengths between 5 minutes and 48 hours were simulated to find the most significant values. The length of the cluster is defined by the last strike inside the time interval, thus most clusters are shorter than their maximum value. The maximum “maximum length” is set to 48 hours to keep a balanced comparison between Correlated and Non-Correlated Clusters. A larger length would add strikes inside the Non-Correlated Clusters defined over the 7 years of the study but not to the Correlated-Clusters.
- Step: the minimum time between the beginnings of two consecutive clusters from a same outage. For simulations, clusters work as time windows which are slid from the step value until another strike is found. For each outage, the first cluster starts with the last correlated strike (the closest in time from the outage reported time), and is then moved from the step value “in the past” and keeps sliding in the past until it finds a new strike. The process is repeated until all the strikes have been attributed to at least one cluster. The step value’s purpose is to prevent the “cluster window” from missing a group of strikes by cutting it in two different parts (clusters). It is possible to ignore the step parameter by choosing its value equal to the maximum length. In this case, clusters are never overlapping. Because all strikes have to belong to at least one cluster, steps values are always smaller or equal than the maximum length of the cluster, even if this causes some strikes to belong to more than one cluster. The step value should not be too small to limit the number of appearance of the same strikes inside different clusters which can cause correlations issues with the statistical model. Thus, the typical step values are the maximum length multiplied by 1/3, 1/2 and 1. It appears that step values have very little impact on statistical model parameters and levels of goodness although steps of half the maximum length tend to give better results in this case. In addition, “1/2-steps” clusters limit the multiple occurrences of strikes to two.

Therefore, for the rest of the study, all clusters used a step value equal to the half of their maximum length.

III. STATISTICAL MODEL

A. Shape of the model

The statistical model used to process the input is the binary logit model:

$$\text{logit}(p_i) = \log(p_i / (1 - p_i)) = k + \beta_1 X_{1n} + \dots + \beta_n X_{in}$$

This model is used because of the definition of the input that uses a time criteria to separate the Correlated and the Non-Correlated strikes. When the model is applied, it is limited to the clusters located inside the Time and Space Window. As a result, the sum of the output corresponds to a “partial sum” and does not represent the real estimation of the number of Confirmed Outages. For this reason, an extra tool is required for the analysis and the tool that seems to work the best is the threshold method proposed here, followed by an aggregation of the models.

The Decision Threshold (DT) of a model is a fixed value above which an observation is considered as a “1”. In other terms, the model gives for each cluster a probability that this cluster actually caused an outage. If this value is above the DT of the model, the cluster is considered as a “1” (caused an outage). The DT of each model is set to maximize the distinction between correlated and non correlated strikes, i.e. the proportion of right answers given by the model on the input data. The model is “correct” if the predicted status of the cluster is the same as the real one.

B. Parameters

In this study, the selection process for parameters is not based on the complex analysis of failure mechanisms. There is extensive literature presenting advanced electromagnetic models in which the goals are to precisely predict the propagation process based on observations made by sensors located near or on the systems. Such a level of precision is not realistic because of the lack of precision of the data available but fortunately, it is also not needed given the statistical point of view of the study. Our approach is more intuitive because it consists of parameters that seem more relevant to the model and have better statistical significance. The first step of the process is the creation and selection of potentially meaningful metrics. Parameters such as the intensity of the strike or the distance between the strike and the system undoubtedly have an impact on the outage mechanism. In addition, the type of system of the area on which it is installed may also play a role. We also developed Hybrid metrics to increase the versatility and precision. Given that the maximum expected precision of the NLDN is 500 meters, all values of distances smaller than that were modified to be equal to 500 meters. The second step is the elimination of non-statistically relevant parameters. First, they have to match the intuitive positive or negative effect they have on the model. For example, intensity should have a positive impact (positive sign for β intensity) meaning that the higher the intensity, the higher the chance for an outage. Then, the significance, represented by the p -value, designates the statistically sound metrics. However, because of important correlation issues between the parameters, the p -value can strongly fluctuate from one model to another for a given parameter.

Table 2: Rejected Parameters

- | | |
|---------------------------|---|
| • Type of equipment | • Highest Intensity/Distance ² |
| • Duration of the Cluster | • Highest Intensity ² /Distance ² |
| • Number of Strikes | • Sum of Intensity/Distance ² |
| • ARTCC | • Sum of Intensity ² /Distance ² |
| • Month / Season | • Duration of the outage |
| • Time Density | |

ARTCC parameters are, however, a special case. Although it appears natural to take into consideration the region where the outage occurs, the definition of the Correlated Strikes already covers this part. It is noted that regions with a larger convective activity see more lightning strikes in their Time & Space Windows. In addition, the characteristics of the strikes relative to a special region (strong short storms, for example) are also covered by the characteristics of the strikes. As a result, taking into account the ARTCC would result in a redundancy, which is why they are not used in the models.

Table 3: Description of the final parameters

Parameter Name	Full Name	Description
H Int over dist	Highest Intensity over Distance	Highest value of Intensity/Distance observed among the strikes of the cluster
H Int	Highest Intensity	Highest value of Intensity observed among the strikes of the cluster
H Dist	Smallest Distance	Smallest value of Distance observed among the strikes of the cluster
S Int over dist	Sum of Intensity over Distance	Sum of all the values of Intensity/Distance of the strikes contained inside the cluster

Table 4: Statistical Model (1hour-long clusters)

Parameter	Estimate	Standard Error	Wald Chi Square	Pr>ChiSq
Intercept	-4.3930	0.1068	1693.178	<.0001
H Int over dist	0.00761	0.000718	112.1786	<.0001
H Int	0.00369	0.00123	9.0080	0.0027
H Dist	-0.3545	0.0419	71.7399	<.0001
S Int over dist	0.000166	0.000037	20.6793	<.0001

The 9 different models are then applied to outage codes different than Cause Code 85-3. First, the Time & Space Windows are created for each outage and then clusters of the different lengths are defined and the corresponding models (with their corresponding codes) are then applied. If, for a given outage and a given cluster length, the models return a value higher than the DT, the outage is considered as a

“Confirmed Outage”. The total number of Confirmed Outages is fairly constant among the different models, but the results for a given outage might vary among models. To consolidate the results, all the models are considered for each outage prediction. Each outage therefore has a “Count” parameter indicating how many models consider it as the result of lightning strikes.

IV. REAL-WORLD CASE STUDY

The model has been developed using 987 outages reported as caused by lightning strikes (Cause Code 85-3) between 1999 and 2005. While building the Time and Space Windows, it appeared that at least 200 Cause Code 85-3 outages did not have a single lightning strike located within 5 NM and 48 hours before the failure. As a result, it is very unlikely that convective weather is responsible for those outages.

The Non-Correlated Strikes were obtained by processing 621 randomly selected outages during 24 months (also randomly selected), which led to a total number of 1,800,000 Non-Correlated Strikes. Then the 9 sets of clusters were created with the different lengths (48 hours, 24 hours, 12 hours, 6 hours, 2 hours, 1 hour, 30 minutes, 15 minutes and 5 minutes) and a binary logit model was computed for each set of clusters. Table 4 is an example of the models for clusters of 1 hour.

In the following step, all the outages contained in the FAA data were filtered to consider only outage codes where lightning strikes could be the original cause (Table 4), and Time and Space Windows were created for them. Unsurprisingly, the correlation levels observed were lower than for the Cause Code 85-3 outages, 27% vs. 80%. Finally, the 9 statistical models were applied to the corresponding sets of clusters and the aggregate method presented above was used. The total number of outages reported under “other codes”, considered by the model as caused by lightning strikes, spanned between 1376 and 2996, which implied a significant modification of the original number.

One of the reasons for the apparent differences between the original data and the output of the model lies in the fact that the FAA code system indicates the ultimate cause of the outage but does not indicate its early cause. For example, for outages reported under “Power Supply” (Code 80-3), nothing indicates what initially caused the loss of power. The first effect tends to overestimate the number of outages caused by lightning strikes whereas the second one underestimates it. However, according to the results of the model, the magnitude of the second effect is larger (plus 1376-2996 versus minus about 200 for the first effect).

Table 7 shows the aggregate results depending on the number of models validating the outages. The most conservative case, where only the outages confirmed by all the models are considered, adds 1376 outages. In this case, even if the new number of outages caused by lightning strikes is almost three times the original one, the corresponding total outage time is only multiplied by two. The systems were separated into 3 categories:

- ATC (Air Traffic Control): regroups all the systems directly impacting the ATC performances: Automated Flight Service Station (AFSS), Airport Surveillance Radar (ASR), Airport Traffic Control Tower (ATCT)
- ILS (Instrument Landing System): contains all the systems reducing the efficiency of the ILS and contains the Localizer, Glide Slope, Markers and also the Lights.
- Other: all other types of facilities

The Confirmed Outages increase the proportion of the ATC and ILS related outages. The proportion also varies significantly with the limit count value while the proportion of the different outage codes under which the outage was originally reported remains almost constant. Future work will focus on what the causes of this phenomenon are, where the proportion of automatic outage report systems and importance of the system to maintain a good level of operations might play a significant role.

Table 4: List of outage codes considered for the estimate

Code	Total Number	Category	Sub Category
80-3	611	Equipment	Power Supply
80-7	13886	Equipment	Unable to Determine Cause
80-F	214	Equipment	Facility Power and Support Systems
81-3	41	Non-FAA Lines/Circuits	Power
81-6	8	Non-FAA Lines/Circuits	Environmental Causes
81-7	99	Non-FAA Lines/Circuits	Unknown
82	4045	Prime Power	-
83	1156	Standby Power	-
84-3	104	Interference Conditions	Radio frequency interference
87-0	6992	Unknown	-

V. CONCLUSION

A correlation between lightning strikes and outages exists and can be captured by statistical tools. To achieve this goal, it is crucial to take into account the important data inaccuracies and to focus more on periods of high convective activity rather than on specific strikes to determine the causes of outages. Both strike localization issues and imprecise outage report times imply a high incertitude in the Target Strikes selection and forced us to introduce the Time and Space Window: a spatiotemporal area of 5 nautical miles and 48 hours associated with each outage and inside which all strikes are considered as a potential cause of outages. Even with such large values, for 22% of the outages reported as caused by lightning strikes (Code 85-3), it was not possible to find a single lightning strike inside the Time and Space Windows. In these cases, it was very likely that these outages were not related with convective weather. It was suggested that such FAA outage codes should probably be modified.

Because of the previous definition of the Time and Space Window, numerous strikes distributed over large periods of time might be responsible for a given outage. In addition, due to randomness of lightning strikes and data imprecision, it was not possible to make a reliable selection of target strikes. For these reasons, we favored an approach based on clusters, gathering lightning strikes over periods of time spanning from 5 minutes to 48 hours, to create independent binary logit models.

Then, we applied models to a selection of codes dealing with power supply failure and unknown causes. The resulting subset contained about 27,000 outages among which 7,300 had at least one lightning strike inside their Time and Space Window. In a number of cases, spanning from 1,376 to 2,996 outages, lightning strikes were likely to be the initial outage cause.

We note that there might be two major sources of inaccuracy that limit the precision of the models. The first and easiest to fix is the localization precision of the strikes which cause the correlation of a significant number of lightning strikes. This problem could be resolved by utilizing a more comprehensive database that would provide the exact location of each stroke, reducing the Space window to an average of 500 meters. The second issue is the outage detection time that can be only partially solved via a list of all the systems

Table 5: Aggregated outages results

		Model Count									85-3 (original)
		1	2	3	4	5	6	7	8	9	
Confirmed Outages		2996	2821	2616	2383	2191	1878	1763	1544	1376	719
Total outage time (1000*hours)		119	102	90	86	83	73	72	65	45	44
Percentage of the total outage time (systems type)	ATC related	7%	8%	8%	8%	7%	7%	6%	5%	7%	4%
	ILS related	47%	51%	56%	57%	58%	58%	58%	64%	50%	45%
	Other	46%	42%	35%	35%	35%	35%	35%	31%	43%	52%
Percentage of the total outage time (reported cause)	Power related	27%	28%	28%	28%	28%	29%	29%	29%	29%	-
	Unknown cause	65%	64%	64%	63%	63%	62%	62%	62%	62%	-
	Other	8%	9%	9%	9%	9%	9%	9%	9%	9%	-

featuring an automatic report device. In these cases, the Time Window would shrink to a few minutes and thus limit the number of Correlated Strikes. In addition, the list of systems using the same power source at airports, and the list of systems featuring forms of protection against lightning strikes, would also improve the overall precision.

The developed methodology and binary logit models should be useful to the FAA management in predicting the number of outages caused by lightning strikes and making appropriate investment decisions regarding the lightning protection and upgrades. These investment decisions are especially important in today's environment, and should ensure that aviation facilities and equipment are additionally protected from any type of convective weather and lightning strikes.

REFERENCES

- [1] E. G. Bates, T. A. Seliga, and T. Weyrauch, "Implementation of grounding, bonding, shielding and power system improvements in FAA systems: reliability assessments of the terminal Doppler weather radar (TDWR)", Digital Avionics Systems, 2001. DASC. The 20th Conference, 14-18 October 2001, vol. 1, pp. 2A5/1 - 2A5/7.
- [2] J. A. Cramer, K. L. Cummins, A. Morris, R. Smith, and T.R. Turner, "Recent upgrades to the U. S. national lightning detection network", 18th International Lightning Detection Conference, Helsinki, Finland, 7-9 June, 2004.
- [3] K. L. Cummins, E. P. Krider, and M.D. Malone, "The U.S. national lightning detection network and applications of cloud-to-ground lightning data by electric power utilities", IEEE Trans. Electromagn. Compat., vol. 40, no 4, November 1998.

Ontology and Rules for International Airspace Security

Mike Renato Henriques
 The MITRE Corporation – Center for
 Advanced Aviation System Development
 McLean, VA, USA
 rhenriques@mitre.org

Abstract- The air transportation system will modernize over the next 15 years. As part of that modernization, tasks that are done manually today will be performed by automated computer functions. In the airspace security domain, automated functions need to be dynamic and able to adapt to the latest intelligence reports. These automated functions can be expressed as “if-then rules.” In order to gain a better understanding of the level of effort involved in creating rules for airspace security, we chose four specific restricted airspaces we felt represented the cadre used to manage security issues and developed the set of terms and relationships needed to define them. This paper outlines the process we took to develop these terms and lists some examples. This paper also presents some sample rules and potential challenges faced in using rules in airspace security. Finally, this paper recommends that in order to obtain a near-term benefit from rules, the airspace security community should consider generalized definitions and broad scopes when developing rules. The potential application and definition of rules is being developed and will be further validated through experimentation in a simulated environment over the next several months.

Keywords: airspace, aviation, security, ontology

I. BACKGROUND

Airspace security is a collaborative activity requiring close cooperation between security partners. Many organizations ranging from military, civil aviation authorities and local law enforcement share the responsibility for identifying, responding to, and mitigating potential airborne threats. On an average day in the United States, that involves air traffic controllers screening over 42,000 flights. And each flight has its own set of characteristics which range from speed and fuel capacity to passengers and crew manifesto, and cargo contents. Each of these pieces of information may be stored in separate databases managed by a different agency. Adding to the challenge, airspace security partners often have ten minutes or less to identify a potential threat and determine the appropriate response.

These challenges can be summed up into three main categories:

- Find the rare event – it can be difficult for humans to sift through a large number of flights and find the one flight that may have significant elevated risk characteristics. This paper does not look at what those risk characteristics might be; however there are efforts within MITRE regarding that topic.
- Manage the “short-fuse” cases – highlighting those cases where the time to reach a valuable target, for example a nuclear power plant, is short.
- Multiple simultaneous incidents – it can be difficult for humans to prioritize events in the case of multiple simultaneous attacks.

The use of customizable rules and computer automation has been suggested as a tool to help the airspace security partners deal with the large quantity of information needed to manage a secure airspace. Rules could allow the personnel to configure which flights are automatically tracked more closely by adjusting the automated rules to fire given the input. This input can be a result of the most recent intelligence report or current activities.

II. INTRODUCTION

The use of rules for automating processes is common in many industries. The insurance industry uses rules to predetermine the level of risk and then assign a monthly premium. Credit bureaus use rules to assign a credit score to those seeking loans, and that affects what interest rates they qualify for. For the purpose of this paper, rules can be thought of as an “if-then” statement. *If* something is true, *then* a conclusion is made or an action takes place.

Although rules architecture can differ from one system to another, the basic notion is that there exists a database of information that is kept up-to-date at a specified frequency. That information is applied against the system rules and feeds into a ‘reasoned’. This ‘reasoner’ is what executes the rules and inputs new information back into the database. The building blocks used to create rules come from ontology.

The definition and use of ontologies – explicit formal specifications of the terms in the domain and relations

among them [3] – has been growing and many disciplines now develop ontologies so that domain experts can use a common and structured vocabulary to share and talk about information in that field. For example, anyone interested in the classification of frogs can go to the Open Biomedical Ontologies website (obofoundry.org) and download ontology for amphibian taxonomy.

There are some existing efforts that provide a structured way of transferring structured data for airspace security. The Aeronautical Information Exchange Model (AIXM) has been developed by the Federal Aviation Administration (FAA) [4] and EUROCONTROL [5] to act as a ‘digital (Notice to Airmen) NOTAM’, with structured information which is suitable for automated computer processing.

The United States (U.S.) Department of Homeland Security has the Homeland Security Infrastructure Protection (HSIP) Gold [6], which is a collection of data and metadata relating to U.S. infrastructure. This collection of data is defined much in the same way as ontology is. This large database includes infrastructure that is relevant to airspace security such as airports, runways, and key locations that may be a target of a September 11 style attack.

This paper uses the restricted airspace domain to develop the specific terms and rules. Restricted airspace is a fixed volume of airspace defined with a start and end time that often prohibits all or most airborne operations. Restricted airspaces are used daily in the U.S. and published on tfr.faa.gov as a tool by the FAA to help manage the airspace from a safety and security standpoint. Some of the instances where restricted airspace can be used for security are around Very Important Persons (VIP) such as a high-ranking government official, sporting events (2010 Olympics in Canada), other high-profile events, and ground assets that might be the target of a 9/11 style attack.

The first part of this paper covers the method used to define the ontology specific to restricted airspace. The results section covers some ontology terms, classes, and some example rules that can be applied to restricted airspace. The conclusion covers what can be done to implement airspace security rules in the near-term. And finally, the future work includes future tasks and contact information.

III. METHOD

Noy and McGuinness’ recommended seven step process was followed to establish a preliminary ontology for dealing with restricted airspace [1]. They are as follows:

1. Determine the domain and scope – decide what will be included and what will be excluded. This is an important first step which generates an ontology that is both effective and reasonable in size. A good ontology is able to describe the things in the scope, and not much more.
2. Consider reusing existing ontologies – there are many aspects of restricted airspace that can reuse existing

ontologies or parts of an existing ontology. For example, distance (radius in nautical miles), time (for the start and end of a temporarily restricted airspace), a point (latitude and longitude in degrees, minutes, and seconds), and so on are all candidates for reuse.

3. Enumerate important terms in the ontology – make a note of the terms that are important to the discussion. What terms could be necessary to use in a discussion with a user?

4. Define the classes and the class hierarchy – decide which terms is either further describing another term or which are generalizations. For example, aircraft could be a top level term, fixed wing and rotor wing are a middle level and 747 and A320 are bottom level terms.

5. Define the properties of classes – each class has different properties that, when put together, form that object. For example, the ‘track’ class – defined as the observed path an aircraft has traveled as noted by returns from radar – associated with the 747 from step 4 has a speed, heading, altitude and beacon code. The ‘flight object’ includes properties such as a aircraft type, nationality, flight plan, tail number, number of armed officers onboard and persons on the watch-list.

6. Define the facets of the classes – classes have different facets that describe things such as value, allowed range of value, and how many values it can have. The class-value is also defined by type. Each slot can be designated as a string (AAL123), number, Boolean (e.g. commercial carrier or not) or enumerated (a choice from a specific list).

7. Create instances – the last step involves choosing a class, creating a specific instance of that class, and defining the facets.

Once the seven steps suggested by Noy and McGuinness were completed, it was then possible to start developing some rules. Using the ontology, sample rules were created that would enforce the airspace restriction and provide a potential benefit to the airspace security partners.

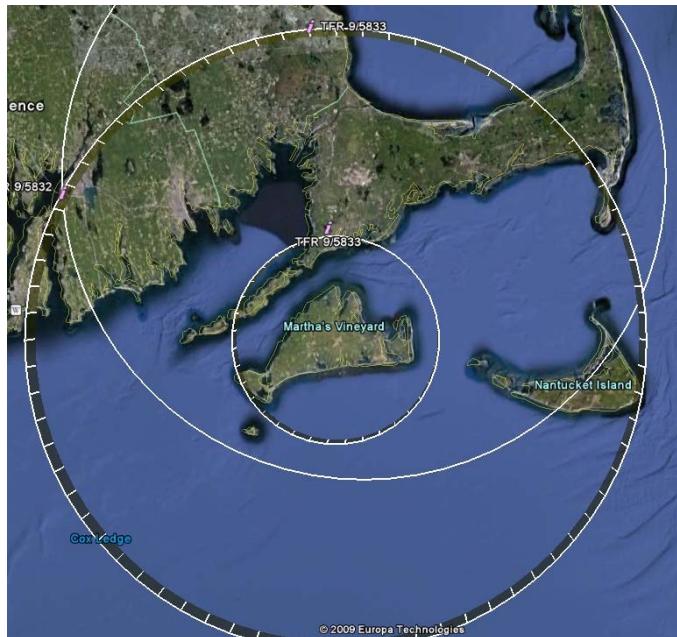


Figure 1. A restricted airspace over Martha's Vineyard, Massachusetts for a VIP involves two concentric circles.
(Map by Google Earth, Google)



Figure 3. A restricted airspace over a California forest fire is constructed using a polygon.
(Map by Google Earth, Google)

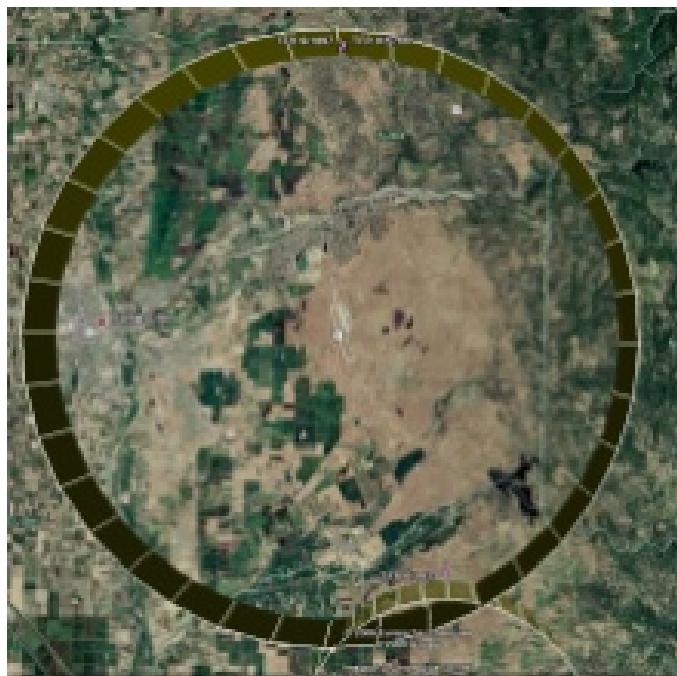


Figure 2. A restricted airspace over Beale Air Force Base is constructed using a cylinder.
(Map by Google Earth, Google)

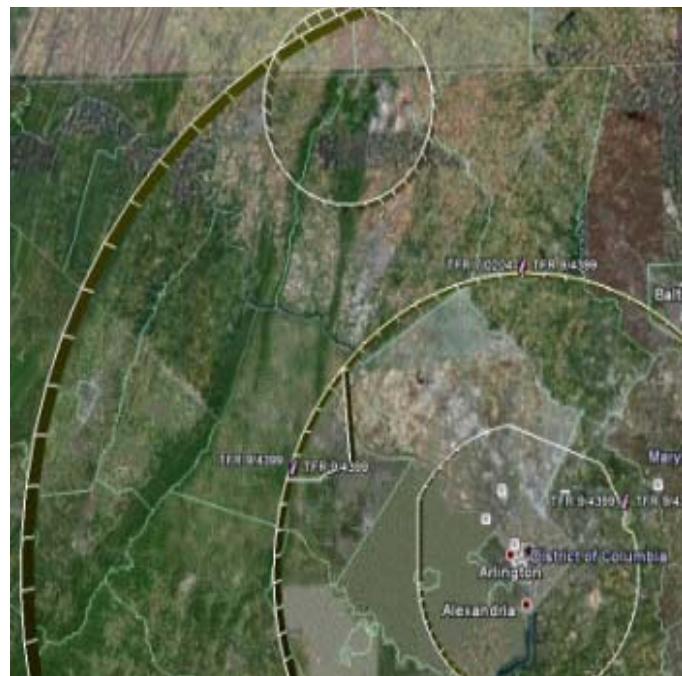


Figure 4. A more complex restricted airspace over the Washington, D.C. area is constructed using a combination of inner and outer cylinders, arcs, planes and cutouts.
(Map by Google Earth, Google)

IV. RESULTS

This section starts by presenting some results we developed from each of the seven steps in section III. Then we present some sample rules created by using the ontology. These rules have not been vetted, but are designed as a first look at what a customizable automated rule could look like for the airspace security partners.

1. Domain and scope – we decided to limit the scope of the ontology to situations dealing with airspace restrictions controlled and managed by the FAA. Not every distinction needs to be accounted for at this stage. For example, knowing which flag the aircraft is flying under may be unnecessary. It may be enough to know whether the aircraft is domestic or foreign.
2. Reusing existing ontologies – unique aspects of restricted airspace definitions were identified that will likely require some modifications to the existing ontologies. In the simple example of time, restricted airspace definitions sometimes have an end time specified as “until further notice,” which is not a standard reference to time. Therefore that term needed to be added. In the case of defining the shape of a restricted airspace, while most are comprised of a cylinder (Fig.1) or a polygon (Fig. 2), some can be hybrids (Fig. 3). These hybrids are combinations of cylinders, arcs and planes. Furthermore, the arcs can be referenced to both clockwise from a point as well as counter-clockwise (in Fig. 4, this can be seen in the inner-most section, between the 2 and 3 o’clock position).
3. Enumerate important terms – as a first step, four FAA restricted airspaces were chosen that represent the sort encountered in the security domain. One was for VIP activity (Fig. 1), one for a forest fire (Fig. 2), one was over Beale Air Force Base where they often conduct Global Hawk Unmanned Aircraft System (UAS) exercises (Fig. 3), and one is over the Washington, D.C. metropolitan area (Fig. 4). In the U.S., restricted airspace definitions are communicated publicly using NOTAMs. NOTAMs are generated by security personnel and are in textual format divided in paragraphs which describe the different restrictions. For example, the inner circle (referred to as ‘Area B’ of the restricted airspace) in Figure 1 was defined as:

Center: MARTHAS VINEYARD VOR/DME (MVY)
(Latitude: 41°23'46"N, Longitude: 70°36'46"W)

Radius: 10 nautical miles

Altitude: From the surface up to but not including 18000 feet MSL

The NOTAM continues on to list the procedures pilots should follow when trying to enter the defined airspace and the restrictions placed on them. The following is an excerpt from the VIP restricted airspace (paragraph A):

All aircraft operations within the 10 NMR area(s) listed above, known as the inner core(s), are prohibited except for: approved law enforcement, military aircraft directly

supporting the United States Secret Service (USSS) and the office of the president of the united states, approved air ambulance flights, and regularly scheduled commercial passenger and all-cargo carriers operating under one of the following TSA-approved standard security programs/procedures: aircraft operator standard security program (AOSSP), full all-cargo aircraft operator standard security program (FACAOSSP), model security program (MSP), twelve five standard security program (TFSSP) all cargo, or all-cargo international security procedure (ACISP) and are arriving into and/or departing from Martha's Vineyard airport (KMVY). All emergency/life saving flight (medical/law enforcement/firefighting) operations must coordinate with ATC prior to their departure at 508-968-7126 to avoid potential delays.

The paragraph begins by stating that all aircraft are prohibited, but then provides a long list of exceptions, which include military aircraft and regularly scheduled commercial passenger flights. The following paragraph from the NOTAM adds an exception for all other aircraft which have applied for and been granted a Transportation Security Administration (TSA) approved waiver (paragraph B):

All other aircraft not operating under a TSA-approved standard security program listed above and arriving KMVY must request a waiver and be security screened at a designated gateway airport. Aircraft departing KMVY during the TFR must also request a waiver and will be screened at KMVY

This waiver is generally used by private pilots who use the airports within the restricted airspace as a destination.

We compared the four definitions; common and differentiating terms used for each definition were extracted. Keeping in mind “what question do we want to be able to answer?” was a key part in going through each restricted airspace definition. An example of such question is “is aircraft X allowed within the boundaries of restricted airspace Y?”

Some examples of the terms generated are:

Geospatial region	Airspace volume
Altitude floor	Altitude ceiling
Center	Radius
Latitude	Longitude
Airport	Natural Hazard
Flight Plan	Transponder Code

4. Define the classes and class hierarchy – the ‘combination development’ process mentioned in the Noy and McGuiness paper was used for this step. [4] Protégé 3.4 was used as the environment to develop our ontology. Developed by Stanford Center for Biomedical Informatics Research, it is free, open-source and is supported by grant LM007885 from the United States National Library of Medicine. First the more salient

concepts were identified and then generalized and specialized them accordingly.

The Protégé software organizes the classes and hierarchy in a graphical format as shown in Fig. 5. All hierarchy starts with ‘thing’ and is then divided into more specific classes. For this example, we show how a geospatial region can be broken down into subclasses until three sub-classes of ‘hazard disaster relief airspace’ is reached. Further sub-division is possible; however this is a suitable level considering the domain and scope of this paper.

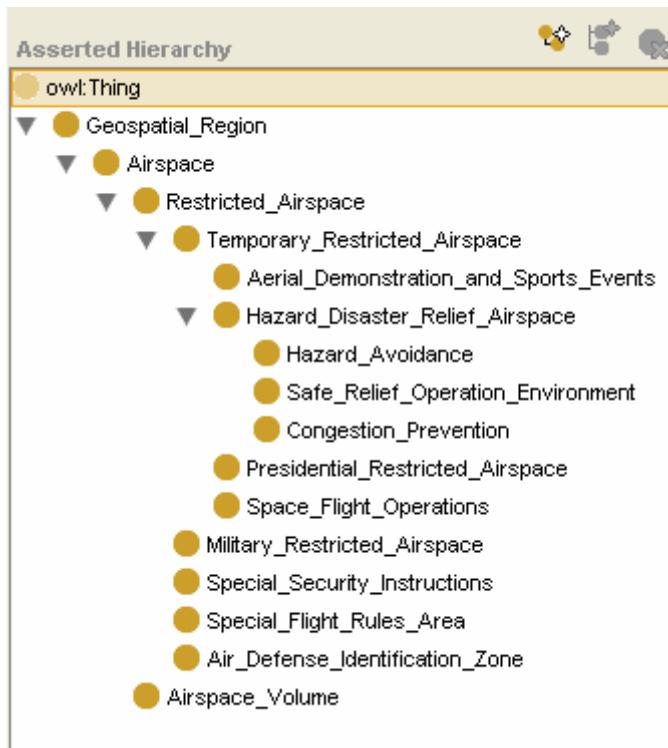


Figure 5. A screenshot from Protégé 3.4 showing the hierarchy for the airspace class.

5. Define the properties of classes – as an example, the properties of a defined airspace are altitude floor, altitude ceiling, center, latitude, longitude, radius, clockwise and counter-clockwise. All four of the restricted airspaces used for this paper can be defined with the use of these properties.

6. Define the facets of the classes – taking the example of a coordinate, the longitude can be any number of degrees between -180 and 180 with minutes and seconds ranging from 0 to 60. A coordinate point used to define a restricted airspace can further be limited to points that lie within the airspace assigned to the country responsible.

7. Create instances – three of the restricted airspace definitions were used in the enumerating terms step (#3) were used as a test for this step. They are the cylindrical restricted airspace over Beale Air Force Base, the polygon over the California forest fire and the hybrid restricted airspace over the Washington, D.C. area. The NOTAM definition of the inner-most portion of the

Washington, D.C. restricted airspace is shown in Figure 1. Fixed Radial Distance (FRD) stands for Fixed Radial Distance)

Upon completion of these seven steps, we had developed a preliminary ontology for restricted airspace that could then be used to define instances. Using the Aero Navigation Aids (Navaid) class as an example; it has a sub-class of VHF Omnidirectional Range (VOR), VOR/Distance Measuring Equipment (DME), VHF Omnidirectional Range Tactical Aircraft Control (VORTAC) and Non-Directional Beacon (NDB). Each of those subclasses has an attribute of a latitude and longitude, which is determined by their geographical location. An instance of a Navaid is the one for John F. Kennedy airport in New York City. Its name is JFK; it is a VOR/DME and has a lat/long of 40-37-58.400N / 073-46-17.000W.

Using permutations of this ontology, many different rules can be constructed, shared and used by security personnel. For example, a user may apply the following rule:

- *If aircraft is inside restricted airspace X and is a banner-towing operation, then sound an alarm.*

Applying this rule to the whole airspace would relieve the airspace security coordinator from manually completing that task.

4-D trajectories forecast where an aircraft will be within a given timeframe and can be helpful to the security coordinator by predicting a violation. It is a computer generated estimation based on the most recent aircraft track. Making use of 4-D trajectories would allow for the use of a rule such as:

- *If aircraft X is headed for restricted airspace Y and it has turned off its transponder, then sound an alarm.*

The ‘is headed for’ portion of that rule could be defined by the security coordinator to be whatever proximity (either in time or distance) he or she feels is best fit to help guard that airspace. Also, the ‘alarm’ action could mean one thing to a restricted airspace over a forest fire (the incursion is likely an accident), and mean something else with a restricted airspace over an inauguration speech (perhaps an attack). In the first case the airspace security coordinator could help alert the pilot, whereas in the second, defense resources may need to be put into place.

In both of these examples, the definition of the restricted airspace can be complicated and involve more terms than the rule itself. The conclusion section of this paper includes a recommendation on how to overcome this.

The following rules can be considered for the enforcement of the VIP restricted airspace referred to in step 3 of section IV. The first line of paragraph A prohibits “all aircraft operations” and would yield in a rule that states:

- *If aircraft is inside the restricted airspace, then alert coordinator.*

However the text continues to explain that there are some exceptions. Military, air ambulance, scheduled commercial and cargo operations for example, are exempt from the restricted airspace. So there is a need to modify the above rule to accommodate this list of exemptions. For example:

- *If aircraft is inside the restricted airspace, and it is not authorized, then alert coordinator.*

The list of what is authorized would represent the specified operations in paragraph A.

Continuing to paragraph B mentions that aircraft that have filed for, and received a waiver from the TSA are allowed to operate within the airspace. Further building on the rule, we get the result of:

- *If aircraft is inside the restricted airspace, and it is not authorized, and it does not have a waiver, then alert coordinator.*

Another requirement we encountered in the NOTAMs is that pilots should monitor the Air Traffic Control (ATC) frequency. We can create a rule for this, such as:

- *If aircraft is inside the restricted airspace and pilot is not monitoring ATC frequency, then alert coordinator.*

This rule might pose a challenge to a rules system since it is difficult to conclude whether or not a pilot is monitoring a radio frequency and is mostly left to the honor system.

Most of these rules are active only while the restricted airspace is active. So a considerable part of developing a rules system would have to consider time and geospatial definitions as an operative.

V. CONCLUSION

The terms necessary to formulate restricted airspace ontology are numerous. The times and irregular shapes are difficult to account for and situations are dynamic. We started with only four NOTAMs to get a sense of the level of effort required to create an automated system for monitoring restricted airspace. Repeating this process for more restricted airspace definitions will be less time consuming, since many of the terms are reused. However, part of the difficulty comes from making the ontology account for rare security events. Today's intelligence report might contain a new watch-list item that was not included in the original ontology definition.

For the airspace security community to obtain the benefits of rules in the near-term, we recommend limiting the scope and complexity of the initial ontology and rules model. For example, instead of listing each specific type of operation not authorized within a restricted airspace (flight training, aerobatic flight, glider operations, parachuting, hang gliding etc...) develop a single category to group operations by risk level. This way, if a new unauthorized operation needs to be addressed, it can be referred to as an existing category. This system is already used in describing in-flight disturbance levels. From 1 to 4, each level carries a

higher impact to the security of the aircraft than the other. Creating rules that monitor aircrafts entering into complex restricted airspaces is another example. These rules need only monitor the general area encompassing the complex shape. This could be accomplished by a single cylinder that assumes most of the complex shape but not all, and covers some areas that are not included by the complex shape. The resulting alarm would bring the approaching threat to the attention of the authority and he/she could monitor that flight more closely.

For this paper, we looked at turning the restricted airspace definitions contained in NOTAMS into rules and the ontology required to support them. Instead of rules being derived from a manually typed up text defining the restricted airspace, a more successful approach might be to develop a user interface that helps define the restricted airspace, its restrictions and that can automatically generate rules. The operations personnel would select from a series of drop-down style computer menus choosing what is and is not allowed.

VI. FUTURE WORK

More work needs to be done to refine and grow the ontology for a rules system model. Looking at more restricted airspaces in the way outlined in this paper will continue to add to the ontology. Eventually the model will approach a complete set of terms required to be useful for operations. That level of usefulness has yet to be determined. At some point, little or no information is added to the model by looking at additional restricted airspace definitions.

The ontology and the rules created for the model should be validated and verified by Subject Matter Experts. This could be done with a prototype rules system and in a simulated environment where SME can assess the rules for effectiveness and the ontology for usability. The simulated environment would also allow for the implementation of test cases to validate rules individually, as well as all together. This simulation would reveal the overlapping of rules and expose gaps in the rules that restricted airspace enforcement.

For more information, please contact the Aviation Security Modernization and Evolution group of MITRE-CAASD at rhenriques@mitre.org.

VII. REFERENCES

- [1] N. F. Noy, and D. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology," Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001.
- [3] T. R. Gruber, A Translation Approach to Portable Ontology Specification, Knowledge Acquisition 5, 1993, 199-220.
- [4] <http://www.faa.gov/AIXM/>
- [5] http://www.eurocontrol.int/aim/public/standard_page/ aixm .html
- [6] http://proceedings.esri.com/library/userconf/fedec08/papers/hifld_hsp_overview_esri_fedec_feb_2008_jms.pdf
- [7] N. F. Noy, and D. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology". Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001.

VIII. DISCLAIMER

The contents of this material reflect the views of the author and/or the Director of the Center for Advanced Aviation System Development, and do not necessarily reflect the views of the Federal Aviation Administration (FAA) or Department of Transportation (DOT). Neither the FAA nor the DOT makes any warranty or guarantee, or promise, expressed or implied, concerning the content or accuracy of the views expressed herein.

©2010 The MITRE Corporation. This document has been approved for public release. Distribution is unlimited.
Case #: 10-1606

Queueing Models for Operations in NextGen

Tasos Nikoleris, Mark Hansen

Department of Civil and Environmental Engineering

University of California at Berkeley

Berkeley, CA, USA

nikoleris@berkeley.edu, mhansen@ce.berkeley.edu

Abstract—This paper develops a queueing model for trajectory-based aircraft operations, a cornerstone of the Next Generation Air Transportation System. Aircraft are assigned scheduled times of arrival at a server, which they meet with some normally distributed stochastic error. A recursive queueing model with deterministic service times is formulated, and Clark's approximation method is employed to estimate each flight's expected queueing delay. The model is further developed to account for aircraft's runway occupancy time, and to track aircraft's delay through a series of servers.

Keywords-queue; aircraft; NextGen; 4D operations

I. INTRODUCTION

The US national airspace system (NAS) is undergoing major transformations, developing towards the so-called Next Generation Air Transportation System (NextGen). NextGen features a shift from the current static system of routes and sectors to one that is adaptive to weather, traffic, and user preferences. System-wide implementation of satellite-based surveillance techniques, primarily Automatic Dependent Surveillance – Broadcast (ADS-B), and navigation methods, such as Area Navigation (RNAV), is expected to greatly reduce human operator workload and significantly increase airport and airspace capacity. Moreover, digital communication links between the aircraft cockpit and air traffic controllers will capacitate information exchange on aircraft's predicted flight path and the negotiation of specific trajectories to be executed. That will allow controlled times of arrival into busy terminals, weather-impacted airspace, and other bottlenecks.

The motivation for this research is the fact that the ability to control and predict 4D aircraft trajectories (4DT) with high precision is a cornerstone of NextGen. 4DT capability, with time being the fourth dimension, is defined as the ability to precisely fly an assigned 3D trajectory while meeting specified timing constraints on arrival at waypoints [1]. This will allow high density flows that rely on controlled times of arrival for critical resources, including entry and exit to/from airspace sectors, taxiways, and runways [1].

However, even with the deployment of the very best 4D trajectory precision and navigation tools, adherence to 4D trajectories will not be perfect. Sources of imprecision include airframe-to-airframe variation in aerodynamic performance, limitations in wind prediction capability, variations in flight crew technique, and varying degrees of exactitude in

navigational performance [1]. As the NAS evolves from its current state to a future condition where location precision is maximized, a spectrum of trajectory uncertainty will be manifested. It will range from low precision, corresponding to today's operations in the NAS, to almost perfect precision, brought on by full deployment of precision navigation and 4DT trajectory awareness tools. For a comparison of delays corresponding to the two ends of this precision spectrum, see [2]. While the models for such cases are well established, it is far more challenging to consider intermediate levels of stochasticity. Such cases are far more representative of the future NAS, in which trajectory adherence will be imperfect. Thus, the objective of this paper is to model 4DT aircraft operations in NextGen using queueing theory, in a way that accounts for levels of trajectory uncertainty in all intermediate phases of precision navigation deployment.

Existing analytical queueing models typically assume that the aircraft arrival process at an airport's terminal airspace area is a non-homogeneous Poisson process [3]. The Poisson-arrivals assumption implies that the variance in total number of arrivals within a given time interval is inherently structured in the model, equal to the mean number of arrivals. However, such a formulation does not control for different levels of uncertainty, which this research study aims to capture by incorporating imprecision in trajectory execution as a parameter in the model. Therefore, a queueing model with arrivals that are scheduled to a server is proposed in this paper, to analyze flight delays in a high-precision trajectory-based operational environment, as currently being planned for NextGen.

Queueing models with scheduled arrivals have been proposed to study port operations. Sabria and Daganzo [4] examine single server queueing systems where customers must be served in an order that is specified by a timetable, i.e. in a First-Scheduled-First-Served (FSFS) manner. Each customer has a scheduled time of arrival at the server, where they actually arrive with some stochastic lateness (positive or negative). Exact transient solutions are obtained for the case when the lateness distribution is Gumbel, and service times are deterministic.

In the present paper, stochastic deviations from scheduled times of arrival are assumed to follow a Normal distribution. Under that condition, exact estimates for each customer's expected queueing delay are intractable. It is, however, feasible

to obtain approximate estimates by employing a well-known technique, the Clark approximation method. We further demonstrate that, in the context of metered aircraft operations, this method yields accurate estimates, when compared to simulation results.

Our analysis begins by focusing at a single server, and without explicitly considering the effect of runway occupancy time in queue propagation. In the second part of the paper, we extend our model to account for aircraft's time to clear the runway, and also to the case with two servers that aircraft must traverse. Those model extensions facilitate the analysis of super-density arrivals in NextGen [1], where aircraft progress through a series of waypoints at controlled times of arrival, on their descent to the runway.

The rest of the paper is organized as follows: Section II presents the general form of our model and discusses the applicability of Clark's approximation method to obtain estimates for the expected queueing delay of each airplane. In Section III the model is further developed to handle aircraft's runway occupancy time as a separate random variable. Moreover, the model is extended to estimate delays when aircraft traverse two consecutive servers. That constitutes the analysis unit for a network of queues in series. Finally, Section IV summarizes our main findings and conclusions.

II. THE MODEL AND AN APPROXIMATE SOLUTION

A. Model Formulation

Our queueing system consists of a single fix, which may be a point in the airspace or a runway's threshold, and of airplanes that must cross it. Aircraft are assigned scheduled times of arrival at the fix, and fly 4D trajectories to traverse it just on time. However, due to imprecise adherence in assigned trajectories, each aircraft's actual time of arrival at the fix has some stochastic deviation from its scheduled time of arrival. The sources of imprecision may include airframe-to-airframe variation in aerodynamic performance, limitations in wind prediction capability, variations in flight crew technique, and varying degrees of exactitude in navigational performance [1]. In addition, consecutive aircraft must maintain a minimum headway h for safety reasons, which can vary over pairs of arriving aircraft. Since air traffic controllers impose the exact values for h , we consider it as a deterministic variable in our model that reflects a particular air traffic control policy initiative. Moreover, we assume that h is the binding constraint among all factors that may affect the required minimum separation between consecutive aircraft.

Following Sabria and Daganzo's approach, each airplane i has an arrival time at the server A_i that consists of a deterministic and a stochastic portion. The deterministic component a_i is the scheduled arrival time at the fix, while the stochastic component is denoted as \tilde{A}_i and represents the lateness (positive or negative) with which the aircraft arrives at the fix, due to imprecision in trajectory adherence. Therefore, we have $A_i = a_i + \tilde{A}_i$.

If deviations \tilde{A}_i 's are small relative to the headway $a_i - a_{i-1}$ between successive scheduled arrivals, serving aircraft on a FSFS order will not result in excessive delays. As an order of magnitude, NextGen planners foresee values of ± 10 seconds for \tilde{A}_i [5]. Under a FSFS queue discipline, the actual time airplane i departs from the server, D_i , would be A_i if there were no queue at the server by the time it arrived, or the time the previous scheduled aircraft $i-1$ crossed the fix plus a minimum required headway $h_{i-1,i}$ between the two aircraft. The actual times that aircraft cross the fix would then be:

$$\begin{aligned} D_1 &= A_1 \\ D_i &= \max(A_i, D_{i-1} + h_{i-1,i}), \quad \forall i \geq 2 \end{aligned}$$

If there were no stochasticity in the system, the deterministic time of departure from the server would be:

$$d_i = \max(a_i, d_{i-1} + h_{i-1,i}), \quad \forall i \geq 2$$

Accounting for stochasticity, the actual departure time from the server of airplane i , D_i , can also be expressed as the sum of a deterministic and a stochastic quantity:

$$D_i = d_i + \tilde{D}_i$$

where

$$\tilde{D}_1 = \tilde{A}_1 \quad (1a)$$

$$\tilde{D}_i = \max(a_i + \tilde{A}_i, d_{i-1} + \tilde{D}_{i-1} + h_{i-1,i}) - d_i, \quad \forall i \geq 2 \quad (1b)$$

We assume that \tilde{A}_i follows a normal distribution with zero mean (without loss of generality), and standard deviation σ_i . Moreover, if \tilde{A}_i 's are correlated, the vector of stochastic errors \tilde{A}_i follows a multivariate normal distribution with zero means, and a covariance structure Σ : $\tilde{A} \sim \text{Normal}(\mathbf{0}, \Sigma)$. The normality assumption stems from the observation that the probability distribution for \tilde{A}_i is generated by convolving the individual distributions of low-correlated stochastic factors. A similar argument is proposed by Meyn and Erzberger [6], who, in a study of scheduling logic and accuracy for terminal area arrival traffic, also approximate the accuracy of flights meeting their scheduled meter fix arrival times with a normal distribution. It should be emphasized, that \tilde{A}_i 's do not represent factors such as severe weather, departure delays, or en-route congestion that cause significant amounts of delays; lateness effects due to such factors have already been incorporated in the calculation of scheduled times of arrival a_i .

In practice, values for standard deviations σ_i could be aggregated to represent classes of aircraft that have similar capabilities of adherence to 4D trajectories. For example, one

could assume two different values for the standard deviation, σ_A and σ_B , in order to roughly represent aircraft with and without Required Navigation Performance (RNP) capability.

B. Solution with Clark's Approximation Method

In (1), for $i = 2$ both terms of the *max* operator are normally distributed. The *max* operation on normal random variables, in contrast to the *add* operation, does not yield a normal random variable. A well-known result due to Clark [7] provides analytical formulas for the mean and variance of the maximum of two normally distributed random variables. Let X and Y be normally distributed random variables, $X \sim N(\mu_X, \sigma_X)$ and $Y \sim N(\mu_Y, \sigma_Y)$, ρ represent the correlation coefficient between X and Y , and Z be the maximum of X and Y , $Z \triangleq \max(X, Y)$. The mean μ_Z and variance σ_Z^2 of Z are then:

$$\begin{aligned}\mu_Z &= \mu_X \Phi(\alpha) + \mu_Y \Phi(-\alpha) + \gamma \varphi(\alpha) \\ \sigma_Z^2 &= (\sigma_X^2 + \mu_X^2) \Phi(\alpha) + (\sigma_Y^2 + \mu_Y^2) \Phi(-\alpha) \\ &\quad + (\mu_X + \mu_Y) \gamma \varphi(\alpha) - \mu_Z^2\end{aligned}$$

where

$$\begin{aligned}\gamma &\triangleq (\sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y)^{1/2} \\ \alpha &\triangleq (\mu_X - \mu_Y) / \gamma \\ \varphi(x) &\triangleq (\sqrt{2\pi})^{-1} \cdot \exp(-x^2 / 2) \\ \Phi(y) &\triangleq \int_{-\infty}^y \varphi(x) dx\end{aligned}$$

The coefficient of linear correlation between Z and a third normal random variable W , $r[Z, W]$, can also be estimated, given that we know the coefficients of linear correlation between X and W ($\rho_{X,W}$), and between Y and W ($\rho_{Y,W}$):

$$r[W, Z] = (\sigma_X \rho_{X,W} \Phi(\alpha) + \sigma_Y \rho_{Y,W} \Phi(-\alpha)) / \sigma_Z$$

The above formulas give the exact mean and variance of Z . The approximation is introduced by assuming that Z follows a normal distribution with mean μ_Z and variance σ_Z^2 . As a result, it becomes feasible to obtain approximate estimates for the moments of the maximum of three or more normal random variables.

In the context of our problem with scheduled aircraft arrivals, Clark's method can be used for all $i \geq 2$ to approximate D_i 's as normal random variables, and estimate their mean $E(D_i)$ and variance $Var(D_i)$ in a recursive manner:

$$E(D_i) = a_i \Phi(\alpha_i) + [E(D_{i-1}) + h_{i-1,i}] \Phi(-\alpha_i) + \gamma_i \varphi(\alpha_i) \quad (2)$$

$$\begin{aligned}Var(D_i) &= (\sigma_i^2 + a_i^2) \Phi(\alpha_i) + \\ &\quad + [Var(D_{i-1}) + [E(D_{i-1}) + h_{i-1,i}]^2] \Phi(-\alpha_i) \quad (3) \\ &\quad + [a_i + E(D_{i-1}) + h_{i-1,i}] \gamma_i \varphi(\alpha_i) - [E(D_i)]^2\end{aligned}$$

$$r[A_{i+1}, D_i] = [\sigma_i \cdot \rho_1 \cdot \Phi(\alpha_i) + \sqrt{Var(D_{i-1})} \cdot \rho_2 \cdot \Phi(-\alpha_i)] / \sqrt{Var(D_i)} \quad (4)$$

where

$$\gamma_i = (\sigma_i^2 + Var(D_{i-1}) - 2 \cdot \rho \cdot \sigma_i \cdot \sqrt{Var(D_{i-1})})^{1/2} \quad (5)$$

$$\alpha_i = (a_i - E(D_{i-1}) - h_{i-1,i}) / \gamma_i \quad (6)$$

and at each iteration i

$$\rho = r[A_i, D_{i-1}], \quad \rho_1 = r[A_{i+1}, A_i], \quad \rho_2 = r[A_{i+1}, D_{i-1}].$$

Note that $r[A_i, D_{i-1}]$ and $r[A_{i+1}, D_{i-1}]$ are obtained through equation (4) in previous iterations. Effectively, the method is implemented by estimating $r[A_i, D_k]$ at each step k , for all $i > k$. Moreover, $r[A_{i+1}, A_i]$ is considered as input from covariance matrix Σ . Equations (2)–(6) are easy to program and they are computationally efficient. Finally, for a stream of N flights scheduled to arrive at a fix, the total expected delay is defined as:

$$E[W_N] \triangleq \left[\sum_{i=1}^N E(D_i) - a_i \right]$$

This completes the formulation of our queueing model. In summary, the model requires as inputs a schedule of arrival times a_i , a capacity profile expressed in terms of required minimum headways $h_{i-1,i}$, and a covariance matrix of trajectory adherence errors Σ . These, coupled with the assumption for normally distributed trajectory adherence errors, enable the estimation of expected flight delays through Clark's approximation method.

C. Approximation Error

Although the maximum Z of two normal random variables X and Y is not normally distributed, our model is based on approximating Z with a normal random variable. In particular, in estimating $D_i = \max(A_i, D_{i-1} + h_{i-1,i})$ it is assumed that D_{i-1} is normally distributed. That enables the estimation of the mean and variance of D_i , which is then also approximated as a normal random variable. However, each pair-wise operation introduces some error that is propagated and might affect the accuracy of our estimates.

To test the accuracy of Clark's Approximation Method in the context of our analysis, several operational scenarios were

considered. The estimates of the analytical queueing model were then compared against the average estimates from 10^4 Monte Carlo simulation runs, which is considered as ground truth.

Each operational scenario was formulated as follows: a total of 120 aircraft must cross a fix, and the minimum required separation between any two successive aircraft is set to $h_{i-1,i} = 30, 60, \text{ or } 90$ seconds. Each aircraft is assigned a scheduled time of arrival at the server $a_i = a_{i-1} + h_{i-1,i} + b$, where b denotes a buffer time inserted. Aircraft arrive at the server with some imprecision that follows a normal distribution and has a standard deviation σ . Zero covariance was assumed across the aircraft arrival times at the server A_i . A total of 90 scenarios were examined:

- 10 different sequences of $h_{i-1,i}$ (each sequence has an equal mix of 30, 60, and 90 seconds)
- $b = 0, 10, \text{ and } 20$ seconds (held constant within each sequence)
- $\sigma = 10$ seconds (uniform across all aircraft), 30 seconds (uniform across all aircraft), and an equal mix of both.

Two metrics for the approximation method accuracy were considered:

- Percentage Error in Total Delay % (PE):

$$\frac{E[W_N]^{appr} - E[W_N]^{sim}}{E[W_N]^{sim}} \cdot 100$$
- Flight Departure Time Mean Absolute Deviation

$$(MAD): \frac{\sum_{i=1}^N |D_i^{appr} - D_i^{sim}|}{N}$$

The first metric evaluates the accuracy of the Clark approximation method in estimating the expected total aircraft delay $E[W_N]$. The second metric evaluates the accuracy of the method in estimating the expected queueing delay for each aircraft.

The results are presented in Table 1. Each entry in the table represents the average value across the ten scenarios of different $h_{i-1,i}$ sequences. The Total Delay PE metric indicates that the approximation method is within -8% accuracy in

estimating the total delay in the system, as compared to simulation. The MAD metric indicates that the approximation method estimates the expected delay of each aircraft with accuracy better than 1 second, on average. The accuracy of the method slightly decreases when the fleet contains aircraft with different navigation capabilities. This must be due to heterogeneity in the variance of the normal distributions for A_i that enters in the \max operator in each step of the recursion. In summary, these experimental results indicate that our proposed model accurately predicts operational consequences of metered operations with good but imperfect 4DT adherence, be expected in NextGen.

III. MODEL EXTENSIONS

A. Runway Occupancy Time (ROT)

So far we have considered a generic minimum separation requirement $h_{i-1,i}$ between two successive arriving aircraft. In this section we distinguish between airborne separation requirement, and the single runway occupancy rule. While the first constraint imposes minimum safety headways between pairs of leading and trailing aircraft when airborne, the second constraint requires that no more than one aircraft may occupy the runway at any time moment.

Similar to the formulation in section II.A, our queueing system consists of a single fix, which is the runway's threshold. Aircraft are assigned scheduled times of arrival at the threshold, which they must cross in the order specified by the schedule. We define as O_i the time period from the moment aircraft i crosses the runway threshold to the moment it has completely exited the runway. Moreover, let $h_{i,i+1}$ denote the required minimum airborne headway at the moment when the leading aircraft i traverses the runway threshold. Letting A_i and D_i be the actual times of arrival and departure, respectively, from the server, we have:

$$D_1 = A_1 \quad (7a)$$

$$D_i = \max(A_i, D_{i-1} + h_{i-1,i}, D_{i-1} + O_{i-1}), \forall i \geq 2 \quad (7b)$$

Therefore, the time when each aircraft traverses the runway threshold is determined by three factors:

- The time it would arrive at the fix in the absence of

TABLE I. RESULTS OF APPROXIMATION ACCURACY TESTS

	Buffer = 0 (sec)		Buffer = 10 (sec)		Buffer = 20 (sec)	
	Total Delay PE	MAD (sec)	Total Delay PE	MAD (sec)	Total Delay PE	MAD (sec)
$\sigma = 10$ (sec)	-0.62%	0.14	-3.26%	0.09	-3.93%	0.08
$\sigma = 30$ (sec)	-0.49%	0.35	-1.69%	0.35	-2.41%	0.31
Mixed	-1.52%	0.89	-5.74%	0.65	-7.70%	0.44

queue, A_i

- The time the previous aircraft crossed the fix plus the minimum required headway, $D_{i-1} + h_{i-1,i}$
- The time the previous aircraft exited the runway, $D_{i-1} + O_{i-1}$.

The shape of the *ROT* distribution may vary among different runways. For example, Xie et al. [8] fit a normal distribution to *ROT* data collected at ATL airport, while Jeddi et al. [9] fit a beta distribution to data from DTW. In this paper, we approximate the probability distribution of O_i as normal and with uniform parameters across all landing aircraft: $O_i \sim \text{Normal}(\mu_O, \sigma_O)$ for all i .

As a result, we can employ the Clark approximation method to estimate the mean and variance of D_i . That is performed in two steps; first we define as $L_i \triangleq \max(D_{i-1} + h_{i-1,i}, D_{i-1} + O_{i-1})$. It can be shown that the coefficient of linear correlation between the two terms in the *max* operator is

$$r[D_{i-1}, D_{i-1} + O_{i-1}] = (\text{Var}[D_{i-1}] / \text{Var}[D_{i-1} + O_{i-1}])^{1/2}.$$

Applying (2)–(6) we compute $E[L_i]$ and $\text{Var}[L_i]$. Next, we use those estimates in the second step to estimate $D_i = \max(A_i, L_i)$, employing again Clark's approximation formulas (2)–(6). Note that A_i is independent of L_i and, as in section II.A, it is assumed normally distributed around a scheduled time of arrival a_i with standard deviation σ_i .

The above model is applicable only when there is evidence that the *ROT* distribution at a given runway can be approximated by a normal distribution. A model with non-normal distribution is the subject of ongoing research.

B. System with two servers

In this section we present a formulation to model the progression of aircraft through a series of servers. That is often the case with operations in the terminal airspace area of large metropolitan airports, where aircraft are metered at entry fixes and must precisely fly an assigned trajectory throughout their descent to the runway. Such procedures are currently in place in PHL [10] and DFW [11], and are expected to predominate in NextGen under super-density arrival/departure operations [1]. We seek to estimate aircraft's expected times of departure from each fix, given scheduled times of arrival at each fix as input.

The following analysis assumes that aircraft are assigned scheduled times of arrival at each fix, and that they cross each fix in the order specified by the schedule. Therefore, it suffices to consider only two fixes, as the extension to three or more is straightforward. Let $D_{i,1}$ denote the time moment aircraft i departs from upstream Fix 1, $D_{i,2}$ the moment when the same

aircraft departs from downstream Fix 2, and F the set of flights that traverse both fixes. Also, let T_i be the unimpeded (from queueing effects) travel time of aircraft i between the two fixes. Consistent with our previous analysis, we assume that T_i 's are normally distributed around t_i 's with covariance structure Σ : $\mathbf{T} \sim \text{Normal}(\mathbf{t}, \boldsymbol{\Sigma})$. The departure time of aircraft i from downstream Fix 2 can be expressed as:

$$D_{i,2} = D_{i,1} + T_i \quad (8a)$$

$$D_{i,2} = \max(D_{i,1} + T_i, D_{i-1,2} + h_{i-1,i}), \quad \forall i \geq 2 \quad (8b)$$

Our goal is to estimate $E[D_{i,2}]$ by employing (2)–(6). The main difficulty arises in (4), estimating the coefficient of linear correlation $r[D_{i,1} + T_i, D_{i-1,2}]$. That is addressed through a series of steps, described in the following algorithm:

Step 0: Estimate $E[D_{i,1}]$ for all aircraft departing from Fix 1 through (2)–(6). For each aircraft's departure time $D_{i,1}$ estimate its coefficient of linear correlation with all preceding aircraft $k < i$:

$$r[D_{i,1}, D_{k,1}] = \frac{\sqrt{\text{Var}(D_{i,1})} \cdot r[D_{i-1,1}, D_{k,1}] \cdot (1 - \Phi(a_i))}{\sqrt{\text{Var}(D_{i,1})}}$$

Step 1: For the first aircraft departing from Fix 2 set

$$r[D_{i,1} + T_i, D_{1,2}] = r[D_{i,1}, D_{1,1}] \text{ for all } i \in F$$

Step k : For all $i \in F$ and $i \geq k$, compute

$$r[D_{i,1} + T_i, D_{k,2}] = [\sqrt{\text{Var}(D_{k,1} + T_k)} \cdot \rho_1 \cdot \Phi(a_k) + \sqrt{\text{Var}(D_{k-1,2})} \cdot \rho_2 \cdot \Phi(-a_k)] / \sqrt{\text{Var}(D_{k,2})}$$

where $\rho_1 = r[D_{i,1} + T_i, D_{k,1} + T_k]$

and $\rho_2 = r[D_{i,1} + T_i, D_{k-1,2}]$.

To estimate ρ_1 , first it can be easily shown that for any pair (i, k) :

$$\text{Cov}[D_{i,1} + T_i, D_{k,1} + T_k] = \text{Cov}[D_{i,1}, D_{k,1}] + \text{Cov}[T_i, T_k].$$

Thus, $\text{Cov}[D_{i,1}, D_{k,1}]$ is computed in Step 0, while $\text{Cov}[T_i, T_k]$ is given as input in Σ . Finally, ρ_2 is computed in step $k-1$.

The reader will recognize that we have outlined a computational procedure for providing estimates of mean

departure times from the downstream Fix 2. Future research will attempt to relax the *FSFS* assumption, to model situations where aircraft approaching from different directions merge at a fix, and adherence to the scheduled order for crossing the fix is not mandatory if an aircraft deviates significantly from its scheduled time of arrival. Moreover, the accuracy of the Clark approximation method in the context of the queueing model with multiple servers needs to be validated against simulation.

IV. SUMMARY AND CONCLUSIONS

In this paper a queueing model for trajectory-based aircraft operations is presented. Flights are assigned scheduled times of arrival at a fix, which they must cross in the order of the schedule. Aircraft meet these times with some stochastic error that is assumed to follow a normal distribution. A recursive queueing model was formulated, and the Clark approximation method was implemented to analytically approximate the mean and variance of individual aircraft delays. The model was extended to include aircraft's runway occupancy time as a separate random variable, and also to capture the progression of aircraft through two servers.

All formulations provide analytical estimates of the expected queueing delay, without requiring any simulation. That, especially for a network of queues, can facilitate the exploration of a wide range of demand and capacity scenarios. Moreover, aircraft precision is handled as a model parameter, thus allowing for sensitivity analysis of delays as a function of adherence to 4DT's. Finally, the accuracy tests presented in this paper indicate that the Clark approximation method can provide with accurate estimates in the context of queueing models with scheduled arrivals at a single server.

ACKNOWLEDGMENT

This research effort was sponsored by NASA under Award # NNX07AP16A. The authors would like to thank Mr. Todd

Farley, AFT Branch Chief at NASA Ames Research Center, for his valuable comments and support.

REFERENCES

- [1] "Concept of Operations for the Next Generation Air Transportation System," Joint Planning and Development Office, Version 2.0, 2007.
- [2] M. Hansen, T. Nikoleris, D. Lovell, K. Vlachou, and A. Odoni, "Use of Queueing Models to Estimate Delay Savings from 4D Trajectory Precision," 8th USA/Europe Air Traffic Management Research and Development Seminar, Napa, CA, 2009.
- [3] B. O. Koopman, "Air-Terminal Queues under Time-Dependent Conditions," *Operations Research*, Vol. 20, No. 6, 1972, pp. 1089-1114.
- [4] F. Sabria and C. F. Daganzo, "Approximate Expressions for Queueing Systems with Scheduled Arrivals and an Established Service Order," *Transportation Science*, Vol. 23, No. 3, 1989, pp. 159-165.
- [5] H. Swenson, R. Barhydt, and M. Landis. Next Generation Air Transportation System (NGATS) Air Traffic Management (ATM)-Airspace Project, Reference Material. External release version, NASA Ames Research Center, June 2006.
- [6] L. A. Meyn and H. Erzberger, "Airport Arrival Capacity Benefits Due to Improved Scheduling Accuracy," AIAA 5th Aviation Technology, Integration and Operations (ATIO) Forum, Arlington, VA, 2005.
- [7] C. E. Clark, "The Greatest of a Finite Set of Random Variables," *Operations Research*, Vol. 9, No. 2, 1961, pp. 145-162.
- [8] Y. Xie, J. Shortle, and G. Donohue, "Runway landing safety analysis: a case study of Atlanta Hartsfield Airport," Proceedings of the 22nd Digital Avionics Systems Conference, Indianapolis, IN, 2003.
- [9] B. Jeddi, J. Shortle, and L. Sherry, "Statistics of the approach process at Detroit Metropolitan Wayne County Airport," Second International Conference on Research in Air Transportation (ICRAT), Belgrade, Serbia and Montenegro, 2006.
- [10] S. Landry, T. Farley, J. Foster, S. Green, T. Hoang, and G. L. Wong, "Distributed scheduling architecture for multi-center time-based metering," AIAA 3rd Aviation Technology, Integration and Operations (ATIO) Forum, Denver, CO, 2003.
- [11] H. N. Swenson, T. Hoang, S. Engeland, D. Vincent, T. Sanders, B. Sanford, and K. Heere, "Design and Operational Evaluation of the Traffic Management Advisor at the Fort Worth Air Route Traffic Control Center," First USA/Europe Air Traffic Management Research and Development Seminar, Saclay, France, 1997.

Applying Economy-wide Modeling to NextGen Benefits Analysis

Katherine Harback, Leonard Wojcik, Jr, Michael B. Callaham,
Shane Martin, Simon Tsao, and Jon Drexler

Center for Advanced Aviation System Development, MITRE Corporation, Mclean, Virginia, United States

Abstract—This paper applies an economy-wide modeling framework, computable general equilibrium, to trace how the Next Generation Air Transportation System (NextGen) could impact non-aviation industries. The specific model used is an adaptation of Monash University's U.S. Applied General Equilibrium model known as USAGE-Air. Modeling results presented here are based on a simple notional representation of NextGen costs and benefits.

Keywords- *economics, investment, NextGen, computable general equilibrium, CGE*

I. INTRODUCTION

Most proponents of National Airspace System (NAS) modernization cite the direct benefits of investing in the Next Generation Air Transportation System (NextGen), including delay reduction, and resource savings from flights using airspace more efficiently, increased system reliability during bad weather, safety improvements, and other potential gains. Many NextGen benefits are only possible with significant investment costs, both public and private. As a result, the Joint Program Development Office (JPDO) and the Federal Aviation Administration (FAA) have been building the case quantitatively to justify the cost of investment to Congress and other stakeholders. Environmental impacts are also being examined for noise, emissions, and climate change. Thus, the NextGen analyses to date have been on impact to the aviation system users and consumers, including airlines, the FAA, the traveling public, consumers of air cargo services or military aviation. (See, e.g., [1]).

Given this industry-level focus on NextGen, additional benefits beyond the aviation industry have not yet been fully studied. The focus of this research is to describe the economic benefits of a notional representation of NextGen to the broader economy outside the aviation industry. This work includes quantifying the impact on other industries in the economy, and describing the impact on such macroeconomic values as Gross Domestic Product (GDP) as well. This puts the NextGen benefits in a broader, economy-wide context and quantifies benefits accruing beyond the aviation industry. It is intended that this research be complementary to ongoing aviation-specific benefits analyses, as the industry level analyses represent and important input for this effort.

While aviation's impact on the economy has been previously assessed [2, 3], this has been from an input-output perspective. There are two general techniques known for capturing quantitative cross-industry impacts: input-output (I-

O) modeling [4] and computable general equilibrium (CGE) modeling [5] (often referred to as applied general equilibrium). I-O models rely on detailed data covering the resource flows between industries. Generally, for the United States (U.S.), these data come from the Bureau of Economic Analysis (BEA). In operation, if an increase in output in one industry is fed into an I-O model, it will produce the increase in gross output due to an increase in the inputs required to produce it. I-O models, however, do not usually include price implications necessary for consistency and reflecting resource constraints. For example, in a pure I-O framework, one might be able to greatly expand a particular industry while never encountering the inhibiting effect of driving up the cost of its material or labor inputs. Chang, et al. [6] compares the prediction of an I-O model to one made using a CGE model. It illustrates that CGE models use the same detailed commodity flow data as I-O models, but add a degree of behavior and dynamics in the form of price responses through demand and supply relationships, and thus achieve a more realistic and consistent result.

I-O analysis (of the past, as distinct from I-O modeling for simulation of possible futures) has been used to estimate the contribution of air transportation, and of the value added by the air transportation industry, to the U.S. GDP in 1992 [2]. The most recent such study [3] looking at the economic contribution of aviation finds that aviation accounts for 5.6% of the total U.S. economy. However, this number does not give us an estimate on the potential broad impact of implementing NextGen. The work presented here using CGE modeling, offering a dynamic capability appropriate for looking forward at how significant changes in the aviation industry could have a broader economic impact.

II. BASIC INDUSTRY ECONOMIC RESPONSES

NextGen will potentially reduce the costs of operating at current levels of traffic through time and fuel saving advances enabled by a variety of technologies, concepts, and capabilities (Required Navigation Performance (RNP)/Area Navigation (RNAV) optimized routes, for example), and by optimizing constrained resources associated with delay. The efficiency gains would also reduce the cost of expanding traffic levels, since the marginal cost of incrementally adding additional flights to a system operating below capacity can be much less than adding flights when demand is already nearly at capacity. This implies an expansion of airline industry supply. This is illustrated in Fig. 1 as the shift from supply curve S₁ to S₂, showing that at any given price, the supplier (the air carrier) would be willing to sell more. Notionally, in Fig. 1, the

quantity is represented as revenue passenger miles (RPMs) and the price could be in terms of yield, fare, or some other form of price.

Moving along the original demand curve, D_1 , this expansion in supply would put downward pressure on prices. Note that demand refers to people buying air transport services from the air carrier (not demand for air traffic services on the part of air carriers). If consumers of air transport services only respond to NextGen by responding to a decrease in prices, their response is captured by the existing D_1 demand curve. However, if enhanced safety, reduced delay, increased reliability, or any other perceived features of the NextGen enhancements prompt them to want more air travel at any given price, then the demand response would reflect an increase in demand, as illustrated by the shift from D_1 to D_2 .

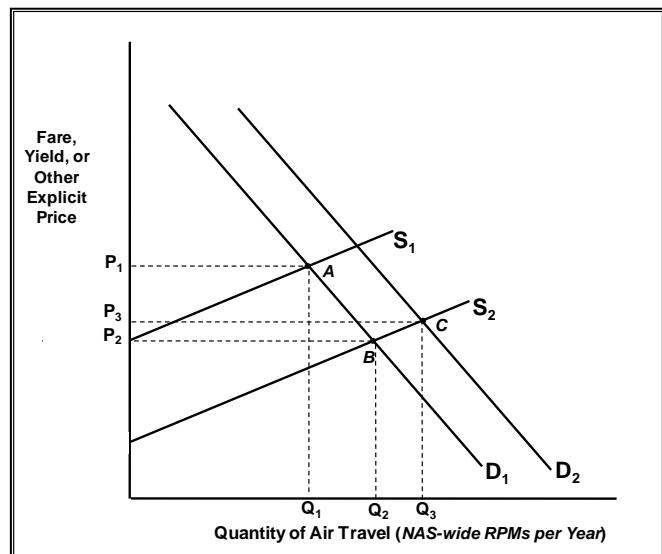


FIGURE 1. NOTIONAL SUPPLY AND DEMAND RESPONSES TO NEXTGEN

With or without the extra demand shift response to NextGen, the amount of air travel expands, as both the increase in supply and increase in demand would put upward pressure on quantity. The increase in supply puts downward pressure on price, while the demand shift, if it takes place, would put upward pressure on prices, making the final impact on prices ambiguous (it would depend on the magnitude of the shifts and the relative slopes of the curves).

III. ECONOMIC RESPONSES—ACROSS THE ECONOMY

To understand how NextGen and the resulting changes to demand and supply inside the aviation market translate into economy-wide impacts, consider Fig. 2. First, NextGen itself would change air carriers' pattern of resource use. This is best illustrated in the case of fuel savings, though relevant to many time dependent resources. An expansion in supply in the aviation industry may also imply that airlines will use more of the resources necessary for producing air transport services. Both of these effects are illustrated as the relationship between the Airline Industry Supply and the Production Demand for Aviation Inputs.

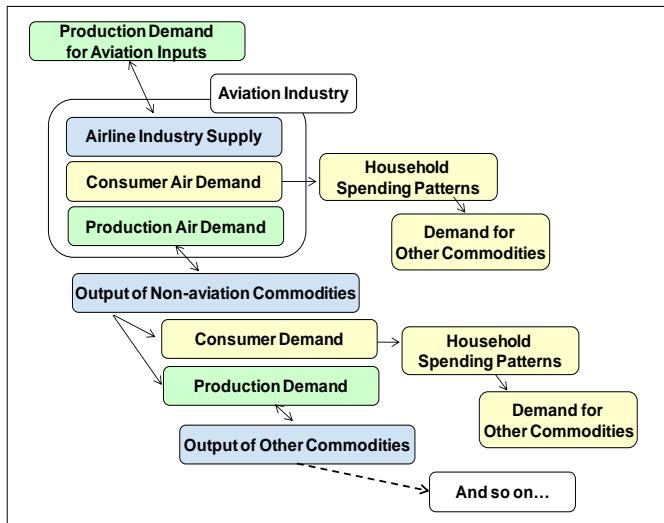


FIGURE 2. THE CONCEPT OF ECONOMY-WIDE LINKAGES

Even if not considering a demand curve shift, falling fares would imply a change in broader household spending patterns, impacting household spending on other commodities. Beyond household demand for air travel, some air travel is consumed by businesses (Production Air Demand in Fig. 2). This could take the form of employees flying to work sites, or movement of inventories, or receipt of shipped supplies. Through this Production Air Demand, the NextGen improvements could have an impact on the output of other commodities, which could cycle through another round of consumer demand impacts and production demand impacts and further.

In most CGE models, e.g. [5, 7, 8], even labor is subject to market clearing. In the short-term, labor employment may be able to rise in the CGE model temporarily until wages adjust (consistent with the macroeconomic concept of “sticky wages”), long run employment levels are considered a function of demographics in the model—markets clear at equilibrium levels and unemployment associated with business cycles is not a focus. This is not a shortcoming of CGE, but rather is reflective of its purpose looking at long run trends and relationships amongst industries rather than modeling business cycles such as the recent recession.

Fig. 3 offers the classic, simple circular flow diagram of the economy. Obviously, the economy is significantly more complicated, but the circular flow diagram captures the big picture of the flows between the consumers and industries, both in the form of demand for final goods and services, but also in the form of labor and other primary factors of production provided by households. Further, Fig. 3 illustrates industries buying outputs amongst each other (outputs traded among industries are called “intermediate” commodities). Foreign trade is a necessary component for consistently capturing all of the economic activity. Finally, Fig. 3 also shows the government related flows in the form of government services and taxes. The circular flow diagram essentially describes the conceptual structure of an economy-wide CGE model.

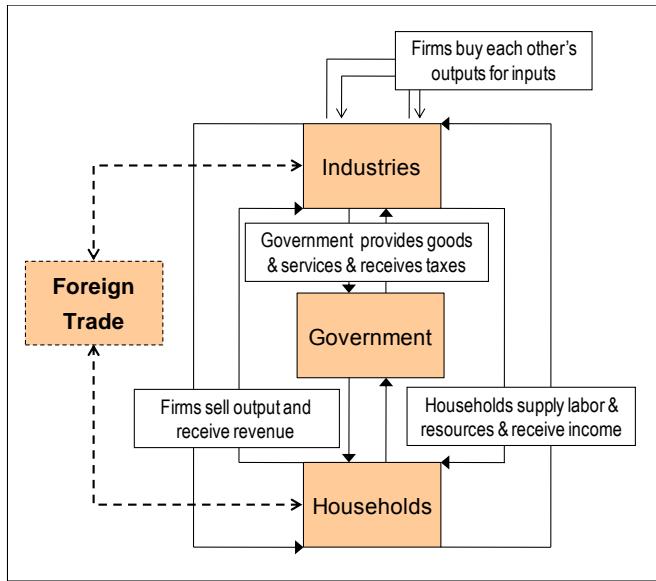


FIGURE 3. SIMPLE SCHEMATIC OF AN ECONOMY OR A CGE MODEL

While there are many CGE models across the academic economic literature, government agencies, and other sources, a smaller number of CGE models are particularly relevant to the economic analyses of U.S. federal agencies. The one that this research focuses on is a variant of the U.S. Applied General Equilibrium Model (USAGE) developed and maintained by the Monash University Centre of Policy Studies [7, 8]. USAGE has been used in analyses for the U.S. International Trade Commission [7], the U.S. Department of Agriculture [9], and other federal agencies. A different model, a version of the Global Trade Analysis Project (GTAP) CGE model, has been used to predict the impact of the U.S.-EU Open Aviation Area Agreement on the economies of the United States and the European Union [10].

The full USAGE model has a level of detail that includes almost 500 industries. There are two more widely circulated “Mini-USAGE” models—one has almost 40 industries and the other [11] only 5, but neither includes an air transport industry distinct from the broader transportation industry, as does the full USAGE model. We have worked with Monash University to bring about “USAGE-Air” [12] which capitalizes on the best, most relevant features of the full and “mini” USAGE versions tailored to analysis of national aviation issues. Currently with 59 industries and 62 commodities, USAGE-Air offers the ease of reasonable run times on a standard personal computer (PC) with a tractable number of variables to work with, but offers significant disaggregation of the air transport industry more like the detail available in the full model. The axis labels of Fig. 6 and Appendix A of [2] have a full list of these industries.

In general, for large, complex model like USAGE-AIR there are two primary types of uncertainty which can affect the validity of the response: 1) structural uncertainty and 2) parametric uncertainty. Incomplete or incorrect knowledge about the relationships, forms, or designs in the system being modeled all contribute to structural uncertainty. The knowledge may include dimensionality, model resolution, missing or poorly understood system dynamics, and

parameterized dynamics. On the other hand, parametric uncertainty is attributable to imperfect information on the values of the inputs or parameters used to calibrate the behavior of a model given a particular model formulation. Parametric uncertainty is expected to be the main vulnerability of USAGE-AIR.

Peter Dixon and Maureen Rimmer have conducted validation of the USAGE model [14] by testing how much and how much better the model can forecast the past, i.e., how the forecast error diminishes, when the model’s input variables which are economic forecasts themselves, are replaced with the actual, historical numbers. Dixon and Rimmer find that in this case the USAGE model’s forecasting performance is greatly enhanced when the model is given the truth about its sets of exogenous variables, macro/energy, trade, and technology/preferences. In the event that the current exogenous variables are the best estimates available, the validity of the model can be improved by understating the magnitude of the parametric uncertainty within the model. This is the approach currently in place for this research, and is continuing to be developed.

IV. REPRESENTING NEXTGEN IN AN ECONOMIC MODEL

Translating operational impacts to inputs for the CGE model runs involves applying quantitative changes to the detailed patterns of resource use for the industry being examined. This pattern of resource use is known as a production function. The production function reflects inputs required to produce output given the state of production technology. Fig. 4 contains a notional representation of this production function relationship, combining labor, facilities, equipment, materials, and services to generate output.

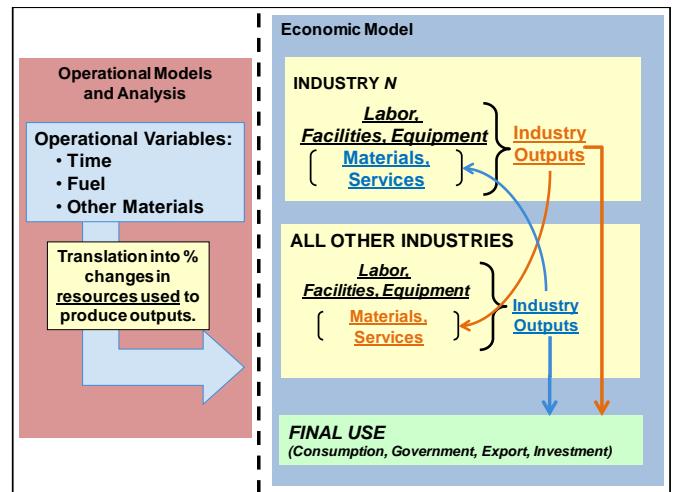


FIGURE 4. PUTTING OPERATIONAL CHANGES INTO THE ECONOMIC MODEL

Fig. 4 also illustrates the relationships between industries’ production functions, as one industry’s output may be used as input to make another kind of good or service. This kind of connection, one industry’s output being used as an input by other industries, is one of the ways effects of NextGen could propagate across the economy. Changes to the production function are applied as percent changes. In some cases,

applying a resource-saving change could result in an absolute increase in resources used. For example, if resources savings result in an expansion of output, an industry could end up using more resources in total. In the model, units of output are measured in 2005 dollars, rather than physical units. These are capable of representing absolute changes in output because changes in prices are tracked in a separate index.

Presently, analyses of NextGen benefits and costs are in a state of being continually updated and refined. Estimates are revised frequently, and are in the process of being vetted and validated. Thinking about NextGen holistically, there is a lot of harmonization of assumptions and pulling together of analyses across different domains and elements to result in a comprehensive benefits and costs portfolio that remains to be finalized. A consequence of this ongoing work is that the results of these studies are not yet published and widely available. For this reason, we chose for these model runs to use a loose, notional representation of NextGen that could be widely discussed. This means our assumptions for the particular model runs presented here are unreferenced, undocumented—we use notional inputs to understand the relationship between critical variables in the output. We are choosing to use these numbers to get a proof of concept that we can put some type of costs and some type of benefits into the model and produce results that will break the ground for the upcoming analysis based on actual, high quality NextGen benefit and cost analysis.

As stated previously, benefits focus on resource savings. While we have the flexibility to input different percentage resource savings for every commodity used by the air transport industry, we chose a single uniform savings for these model runs. In other words, after NextGen is implemented in our notional case, the reduction in delays and overall improvements in efficiency imply 9% fewer resources are consumed to produce any given level of output. This number chosen while awaiting the comprehensive benefits studies we will exercise the model on eventually. While there could be significant difference between fuel savings and the saving of block time sensitive resources, such detailed specifications of benefits will be the subject of futures studies carried out in closer coordination with ongoing NextGen studies. Explicit government resources savings (air traffic management or otherwise) are not included in this analysis.

Significant investment will be required to make NextGen a reality. This will include equipage by operators as well as investment in ATM on the part of the FAA and the government. We assumed a notional cost of \$40 billion dollars. A more detailed estimation may suggest higher or lower costs and will no doubt depend on which user groups are targeted for equipage and when. Given the \$40 billion estimate, we assume that this will be distributed as approximately a 50-50 split between aviation industry investment in the form of equipage and government investment in air traffic infrastructure.

In the model, investment takes place to build capital, which is a feature in the production function. Here, capital refers to the types of physical resources required to produce output that are not entirely consumed in the production of the output—

meaning they last for continued reuse over time, with some amount of depreciation (or wearing out) as they are used over time. Investment decisions are based on expected earning associated with buying more capital (expected rate of return), the existing current capital stock, and the rate of depreciation. To achieve the NextGen investments in our scenarios, we alter the amount of navigation equipment in the investment profile for the government and/or aviation industry. The percents associated with the increased navigation equipment investment were chosen to result in approximately \$40 billion in real spending (as opposed to nominal, which would include inflation). Achieving the cost target of \$40 billion in the model required “tuning” a percentage of increased navigation equipment that results in about \$40 billion in spending.

While it is notable that we have communications and computer-related industries that could be the source of potential NextGen components present in the model, at the economy-level these industries are dominated by things like fiber optic cable, telephone service, and consumer computer products. It is an area being investigated based in the aggregation of the data how much of the actual NextGen infrastructure will come from each industry—while the flexibility exists to slice it across many, until we have a specific portfolio and that additional detail on industry aggregation relative to the portfolio, we kept the scenarios simple and based in the navigation equipment industry.

In our simple notional scenarios, we assumed NextGen implementation, for both benefits and investment costs, would begin in 2010 and concluded by 2025. This is not based on a specific portfolio, implementation concept, or anything other than the basic potential milestone dates that have been broadly discussed for NextGen. Benefits were assumed to accelerate (“ramp up”) continuously over this period to reach 9% in 2025. Investment costs ramp up from 2010 to 2014 to reach a steady level maintained from 2015 to 2023 and then decline to zero across 2024 and 2025 (“ramp down”).

While our comprehensive model runs for this phase of our work consisted of 35 different scenarios, for this paper, we present two scenarios for NextGen implementation. The 35 scenarios represented different range of variation on the basic notional assumptions, and included some model runs touching on subsidies and fees. The difference between the two scenarios presented here is in who bears the \$40 billion in investment costs. In the first scenario the carriers and the government are financially responsible for their own direct portions of the NextGen investment (meaning carriers pay for equipage in the previously discussed 50-50 split) and in the second scenario, the government pays for the entire investment through deficit spending.

In addition to these two scenarios runs, a base case is also run. The base case represents the course of the economy absent any of the NextGen changes. It draws on growth forecasts from the Congressional Budget Office, U.S. Department of Energy, Energy Information Administration, U.S. Department of Labor Bureau of Labor Statistics, U.S. Department of Agriculture, as well as trends in historic data describing consumer preferences, technology, world demand for U.S. exports and U.S. demand for imports [7, 8, 12]. The

focus of this base case forecast and application of the model in general is for understanding the overall, long run, equilibrium movement in the economy (equilibrium characterized consistent with the economic definition—market clearing). This means that we do not address business cycles—including the present recession. It is not that this is not an important focus of economic analysis in general, nor is it an insufficiency of this work—it is just not a feature of the questions this research addresses.

V. ECONOMY-WIDE NEXTGEN IMPACT

Fig. 5 describes the high-level output from the two scenarios. These results are presented in terms of cumulative percent deviation from the base case.

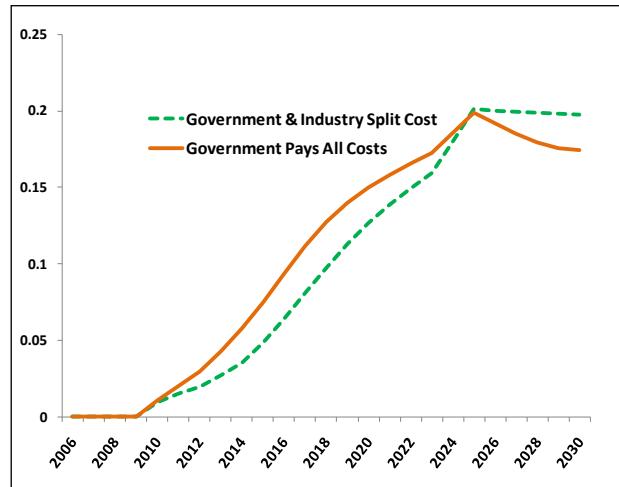


FIGURE 5. PERCENT INCREASE IN GROSS DOMESTIC PRODUCT ASSOCIATED WITH EACH NOTIONAL SCENARIO.*

*NOTE: 0.01 ON THE VERTICAL AXIS IN FIGURE 5 IMPLIES 0.01%

The specific variable presented in Fig. 5 is GDP. GDP is a measurement of the final output of the economy, not counting goods and services used as intermediates to produce other goods (no double counting). It essentially represents the amount of economic output available to support an economy's standard of living. While the impact on GDP is measured in fractions of a percent, the magnitude of GDP means that these fractions translate to large absolute amounts. The approximate level of both scenarios in 2025 is \$46 billion 2005 dollars.

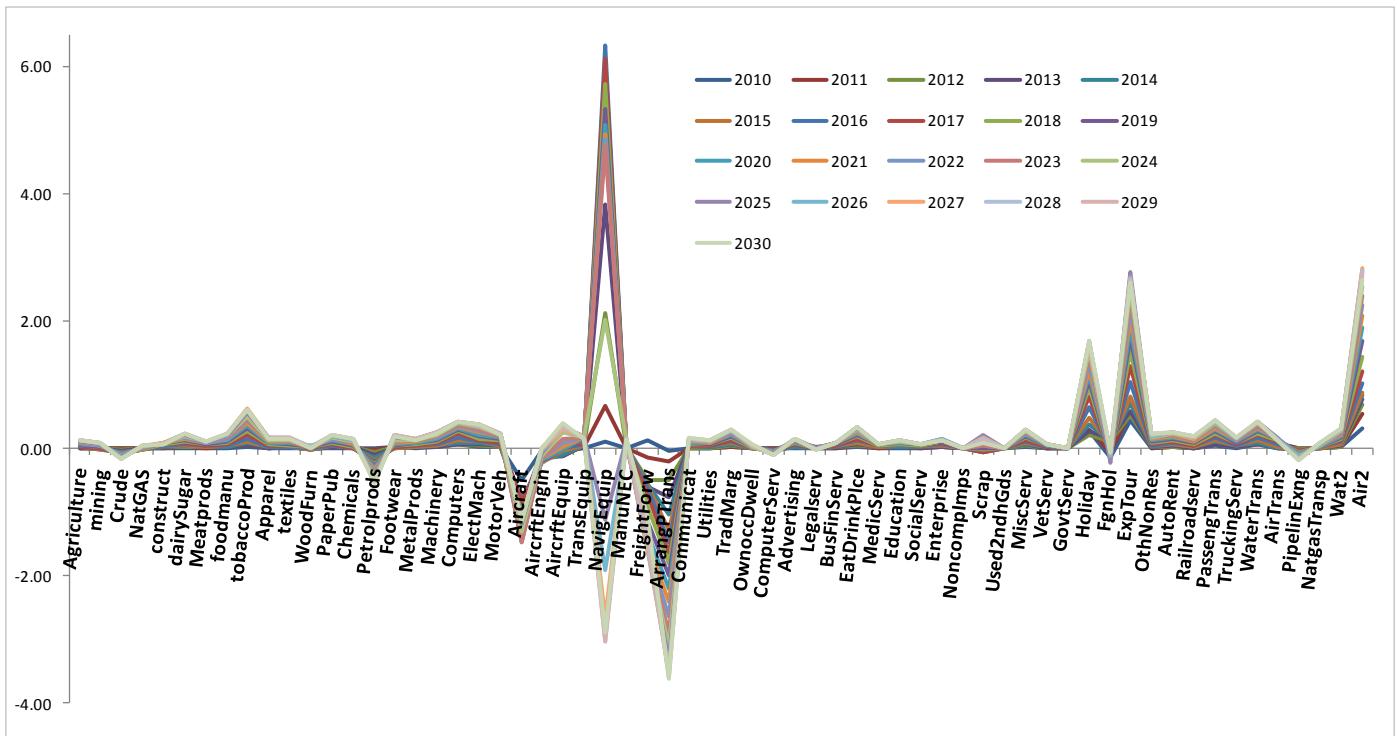
It is notable that both scenarios result in about one-fifth of a percent increase in GDP by 2025. This seems to indicate that at the broad economy-level, it makes little difference in the long run whether the government pays for NextGen or the carriers do, though in the mid-term, there is a difference in the path to that result. After converging at 2025, the scenarios diverge moving forward toward 2030. The industry and government split case is high after 2025 relative to the government only scenario. This is driven by the investment dynamics and the model's equilibrium characteristics with respect to labor, capital, and their respective factor payments. The investment pattern is shocked over a period of time changing the relationship between capital and labor [2]. When the investment shocks ramp down in 2025, there is a temporary adjustment period to achieve the long run equilibrium that

results in a fleeting boost to GDP. It is expected that running these models out further into the future would achieve greater convergence in the cumulative GDP impact.

Despite this issue with the investment costs, separate cost only and benefits only model runs reveal that both scenarios are primarily dominated by the benefits aspect of the scenarios. Understanding the smaller impact of the cost requires understanding the difference between industry-level analysis and economy-wide analysis. In an industry-level study, the \$40 billion in assumed investment costs counts as a whole. When examining GDP changes here, the \$40 billion of extra spending on navigation equipment represents a deviation of resources and government debt (depending on the scenario). Unlike a direct cost benefit analysis, this \$40 billion does not represent a negative in its entire amount, because it still adds up to being part of GDP. This can be seen by examining the navigation equipment industry in the industry-level results. The deviation in resources from their baseline state which implied profit maximization/cost minimization does have some cost, reducing GDP growth.

Fig. 6 is a plot of the specific industries included in the model with their cumulative percent deviation in output (gross output, including outputs used as intermediates) year-by-year for the scenario in which government and industry share the \$40 billion of investment. The industries are roughly organized in the plot by their type—starting with agriculture and natural resource-oriented industries to the right, then manufacturing toward the middle and services toward the left.

Navigation equipment is the industry directly driven by the investment cost assumptions in the two scenarios. In the years between 2010 and 2025, the level of investment in navigation equipment is pushed artificially high relative to the base case. Once the investments are complete, there is a recoiling to below the original investment level for years like 2026 reflecting the larger amount of relatively new navigation equipment in the capital stock. This is a possible area for tuning the scenario inputs to be more realistic, if we think navigation equipment will permanently represent an increased share of the air transport industry's investment profile. The increase in air transport ("AirTrans") appears small, given this is the industry directly impacted by the resources savings in the scenario. Investigation into the parameterization of the model reveals the likely cause of this—the own price elasticity of demand here for AirTrans is -0.8. This is inelastic, meaning for every one percent the price of AirTrans falls, there is about eight-tenths of a percent increase for AirTrans demanded by consumers. There is broad literature that estimates and describes the elasticity of demand for air transport, a great deal of which is summarized in Gillen, et al. (2002) [15]. Of the studies summarized and reviewed by Gillen, elasticity has been estimated as elastic as -3.2, with a median of -1.22. These are both elastic (less than -1) implying greater than proportional response for price changes—unlike -0.8. The choice of elasticity in the current specification of USAGE Air was driven by Monash's broad parameterization of the whole economy, as this was a parameter calculated in common with other modes of transportation. Future model runs will use elasticity estimates from the literature and include sensitivity analysis.

FIGURE 6. PERCENTAGE DEVIATION FROM BASE INDUSTRY OUTPUT IN THE GOVERNMENT & INDUSTRY INVESTMENT NOTIONAL SCENARIO.*

*NOTE: 1 ON THE VERTICAL AXIS IN FIGURE 6 IMPLIES 1%

It should be noted that demand response would result in a larger air transport increase (see the discussion of the shift from D1 to D2 in relation to Fig. 1).

While the increase in air transportation industry was mild, the increase in domestically produced Air2, the industry that represents international flights by U.S. flag carriers experiences a much larger increase in output. Air2's elasticity is -1.5 (meaning a 1.5% increase in quantity demanded for a 1% decrease in price). This effect may be strengthened by Air2 being subject to foreign competition (resource savings make them relatively more competitive and divert some international traffic away from foreign carriers).

Other spikes that might seem unusual are also present—for instance, aircraft. When we introduce blanket resource savings to the air transport industry, we see a small increase in air transport output relative to those resource savings, so resources purchased in aggregate actually decline—producing a reduction in output in the aircraft industry. This is also observed for petroleum products and other supporting industries as well. Inclusion of a demand response, described above, could result in an absolute increase in absolute resource consumption.

The “arranging passenger transport” industry experiences a large decline. It sounds like an industry that should experience a benefit from resource savings in the air transport industry, but arranging passenger transport is an industry that includes tour operation and tours excluding sightseeing by various modes, travel agencies, carpool and vanpool arrangement, and ticket offices not operated by transportation companies. About 90% of the output of the arranging passenger transport industry is

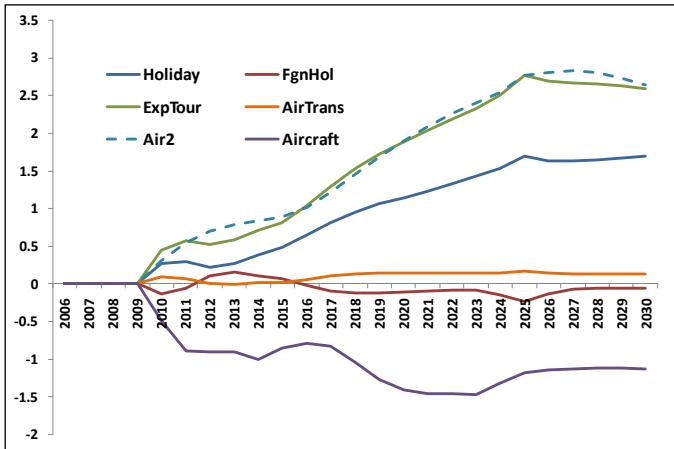
consumed as intermediates—with 40% total being used as input to air transport itself. As described for the case of aircraft, with the specification of these scenarios, industries that produce resources used by air transport experience a decline in output associated with the specified resource savings. This is also the case for freight forwarding, which includes different non-air freight, courier, and warehousing services, though to a lesser degree. Both freight forwarding and arranging passenger transport are relatively small in absolute magnitude. Table 2 contains their base value in millions of 2005 dollars along with the magnitude of some other industries of interest.

Holiday, foreign holiday (FgnHol), and export tourism (ExpTour) are three further industries of note included in Table 2. Holiday, otherwise known as vacation, represents tourism. Foreign holiday is tourism by U.S. nationals abroad. Export tourism describes tourism by foreign nationals in the U.S., which, because it represents foreign purchase of U.S. goods, counts as an export. These industries have no capital and represent bundled consumption that includes air transport, thus the growth of holiday and export tourism. Foreign holiday experiences a slight decline—which makes sense as we have made domestic tourism more competitive to the foreign and the improvement in productivity leads to favorable terms of trade effect for US goods. Fig. 7 is a time series plot of the tourism industries, air transport, Air2, and aircraft (as representative of industries that produce air transport inputs). This is the same data plotted in Fig. 6.

TABLE 2. 2005 GROSS OUTPUT IN MILLIONS OF 2005 DOLLARS

Industry	2005 Output
Holiday: An industry with no capital that combines other industries outputs as domestic vacation spending by U.S. travelers	\$360,990
Trucking Services	\$330,261
Export Tourism: Like Holiday, except that it is consumed by foreign travelers to the United States and is thus an export	\$192,736
Air Transportation: Domestic air transportation services including passenger and cargo	\$134,509
Foreign Holiday: Like Holiday, but consumed abroad by U.S. travelers	\$85,397
Railroad Services	\$79,018
Aircraft Manufacturing	\$65,558
Air2: Air transportation to/from foreign destinations by U.S. carriers	\$52,233
Passenger Transport: Local and intercity passenger transport including taxi, bus, bus charter	\$42,631
Water Transportation	\$41,706
Navigation Equipment	\$37,051
Freight Forwarding: Freight forwarding, warehousing & storage except by air, including local trucking and courier services	\$26,903
Arrangement Passenger Transportation: Includes travel agencies, ticket offices not operated by transportation companies	\$25,142

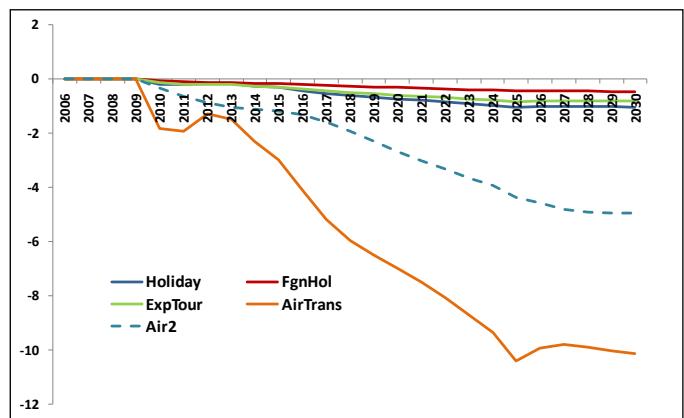
It is even more apparent in Fig. 7 that the deviation from the base for air transport is modest compared to that of holiday and export tourism. This is easily explained in the case of export tourism—it tracks very closely with Air2, the variable describing international flights carried out by U.S. carriers coming or going from the U.S. Explaining the difference in the apparent impact on output for holiday and air transport requires looking at the deviation in the price level from the base case associated with the scenario. These price deviations are presented in Fig. 8.

FIGURE 7. INDUSTRY OUTPUT DETAIL IN THE BENEFITS & COSTS (GOVERNMENT & INDUSTRY) NOTIONAL SCENARIO*

*NOTE: 1 ON THE VERTICAL AXIS IN FIGURE 7 IMPLIES 1%.

Air transport is about 11% of total intermediate goods that go into holiday. However, as holiday is an input to export tourism, there is also some Air2 in holiday. The decrease in

price in air transport, the price of an input, would stimulate an increase in supply of holiday, as would the decrease in price of Air2. While the proportional increases in air transport output may not seem large enough to support the expansion of holiday, understanding that Air2 is also an input to holiday as holiday is an input to export tourism helps reconcile this.

FIGURE 8. CUMULATIVE PERCENT DEVIATION IN PRICE LEVELS IN THE NOTIONAL SCENARIO FROM THE BASE*

*NOTE: 1 ON THE VERTICAL AXIS IN FIGURE 8 IMPLIES 1%.

The price decline in air transport is dramatic. A decline in price and increase in output are consistent with the movement from equilibrium point A to equilibrium point B in Fig. 1. The relative magnitude of the price relative to the increase in output implies a relatively steep demand curve—consistent in principle though not magnitude with the elasticity assumptions described previously. This means that further dynamics in the model are at play beyond multiplying an own price elasticity parameter by an estimated price change. The source of part of these dynamics is the relationship between Air2 and air transport, though a comprehensive view of the dynamics is still be traced out and described.

VI. SUMMARY: THE ECONOMY & NEXTGEN

This paper provides a first look at the impact of investment in aviation infrastructure on the rest of the U.S. economy. A number of extensions will make this model stronger over the coming months. For instance, the model is calibrated to an economic “status quo” with gradual efficiency gain over time in the production functions. If the amount of resources required for air transport in the NAS is increasing with the level of NAS congestion (i.e., the amount of time, fuel, etc. required increases per flight as congestion mounts, a.k.a. decreasing returns to scale), the current production function understates the excess congestion costs of “do nothing” future base case, absent NextGen. This potentially understates the benefits of the NextGen resource savings. If passengers respond directly to reduced delay, increased reliability, and improved safety, as described by D2 in Fig. 1, results could also be more dramatic. If businesses redesign their distribution networks, relying more heavily on air transport to deliver their goods to market, the broader economic impact would also be altered. Finally, resource savings to the government are also not considered in these model runs.

These major refinements are all underway. All have a tendency to intensify the value of NextGen in the analysis. This work is a good example of modeling NextGen-style benefits and costs with CGE modeling, showing it is capable of estimating the value to the economy of investment in air traffic modernization. The economy-wide benefits in terms of gain in GDP strongly dominate the costs in the scenarios presented here. Additional scenarios not groomed for this paper, demonstrate similar features and a strong, positive economic impact for NextGen, given our notional assumptions.

GDP is a modest proxy for the actual improvement in well being. It has the advantage of being a familiar concept with a clear explanation, but as measured in models like these, it can be subject to index problems and does not necessarily convey the entire magnitude of the impact. Compensating variation (CV) and equivalent variation (EV) are two concepts in economics used to more thoroughly describe welfare improvements or deteriorations. Compensating variation roughly describes how much money could be taken away from households to leave them just as well off as in the base case, given the policy scenario (assuming a beneficial scenario—given a negative scenario money would be given). Equivalent variation is roughly how much money would have to be given to households in the base case to make them as well off as they would have been in the policy scenario (assuming a beneficial scenario). These concepts are very similar but not the same [5, 16]. In the scenario for industry and government sharing the cost of NextGen, the model estimates the upper bound on equivalent variation at 1.32% of household expenditures, and the lower bound of compensating variation at 1.29% of household expenditures in 2025.

Efficiency gains in the air transport industry, such as those associated with NextGen, could be great. These benefits appear only modestly diminished when modeled with their investment costs, whether the costs are borne jointly by the industry and the government or solely by the government, given the assumptions made for this application. Whether measured in terms of GDP (0.2% higher GDP in 2025 translates to \$46 billion in 2005 dollars) or with the technical economic welfare measures like equivalent variation and compensating variation (1.32% and 1.29% of household expenditures, respectively) our scenarios demonstrate the value of applying CGE modeling to capture NextGen benefits across the economy.

ACKNOWLEDGMENTS

We offer special thanks to our colleagues George Solomos, Debra Pool, Glenn Roberts, Felipe Moreno-Hines, Joseph Sinnott, and Gene Lin for their criticism, encouragement, review, and support in pursuit of this work.

REFERENCES

- [1] Borener, S., Economic Performance and NGATS, NEXTOR 2nd National Airspace System Infrastructure Management Conference, 2006, [http://nextor.org/Conferences/200606_Infrastructure_Management/2006_06_13_4A_Borener_Alt.pdf].
- [2] Callaham, M., 1999, "Trickle Down: Whence U.S. Air Transportation Industry Revenues Come, and Where They Go," pp. 593-609 in Butler, Gail F., and Keller, Martin R., eds., *Handbook of Airline Finance*, First Edition, (New York, NY: McGraw-Hill).
- [3] FAA Air Traffic Organization, *The Economic Impact of Civil Aviation on the U.S. Economy*, December 2009. [http://www.faa.gov/air_traffic/publications/media/FAA_Economic_Impact_Rpt_2009.pdf]
- [4] Miller, R.E., and P.D. Blair, *Input-Output Analysis: Foundations and Extensions*, 2nd ed., 2009, New York: Cambridge University Press.
- [5] Dixon, P. and M. Rimmer, *Dynamic General Equilibrium Modeling for Forecasting and Policy*, Elsevier, Amsterdam, 2002.
- [6] Chang, C., et al, "An Economy-wide Analysis of Impacts of Avian Flu Pandemic on Taiwan," 2009. [<https://www.gtap.agecon.purdue.edu/resources/download/4278.pdf>]
- [7] United States International Trade Commission, *The Economic Effects of Significant U.S. Import Restraints*, Fifth Update 2007, Publication 906, [<http://www.usitc.gov/publications/docs/pubs/332/pub3906.pdf>]
- [8] Dixon, et al, USAGE Technical Documentation [<http://www.monash.edu.au/policy/usage.htm>].
- [9] Gehlhar, M. and E. Dohlman, Nora Brooks, Alberto Jerardo, and Thomas Vollrath, *Global Growth, Macroeconomic Change, and U.S. Agricultural Trade*, Economic Research Report No. (ERR-46) 44 pp, September 2007. [<http://www.ers.usda.gov/publications/err46/>]
- [10] Peterson, E., and T. Graham, "Open Skies: An Assessment of the US-EU Open Aviation Area Agreement," 11th Annual Conference on Global Economic Analysis, Helsinki, Finland, 2008. [<https://www.gtap.agecon.purdue.edu/resources/download/3848.pdf>]
- [11] Dixon, P. and M. Rimmer, "Mini-USAGE: Reducing Barriers to Entry in Dynamic CGE Modeling," presented at the Annual Conference on Global Economic Analysis, 2005. [<http://www.monash.edu.au/policy/ftp/miniusage/minuse3.pdf>]
- [12] Dixon, P. and M. Rimmer, *USAGE-AIR: a general equilibrium model of the U.S. economy with emphasis on the Air Transport industry*, unpublished, March 2009, updated June 2009.
- [13] Harback, Katherine, et al., 2009, "Bringing an Economy-wide Perspective to NextGen Benefits Analysis," 9th AIAA Aviation Technology, Integration, and Operations Conference, AIAA paper AIAA-2009-7061 [<http://www.aiaa.org/>].
- [14] Dixon, P. and M. Rimmer, *Validating a Detailed, Dynamic CGE model*, 2009. [<http://www.monash.edu.au/policy/ftp/techusage3.pdf>]
- [15] Gillen, Morrison, and Stewart, *Air Travel Demand Elasticities: Concepts, Issues and Measurement*, 2002. [http://www.fin.gc.ca/consultresp/Airtravel/airtravStdty_eng.asp]
- [16] Mas-Colell, Andreu, Michael Whinston and Jerry Green, *Microeconomic Theory*, Oxford University Press, New York, 1995.

DISCLAIMER

The contents of this material reflect the views of the author and/or the Director of the Center for Advanced Aviation System Development, and do not necessarily reflect the views of the Federal Aviation Administration (FAA) or the Department of Transportation (DOT). Neither the FAA nor the DOT makes any warranty or guarantee, or promise, expressed or implied, concerning the content or accuracy of the views expressed herein.

This is the copyright work of The MITRE Corporation and was produced for the U.S. Government under Contract Number DTFA01-01-C-00001 and is subject to Federal Aviation Administration Acquisition Management System Clause 3.5-13, Rights in Data-General, Alt. III and Alt. IV (Oct. 1996). No other use other than that granted to the U.S. Government, or to those acting on behalf of the U.S. Government, under that Clause is authorized without the express written permission of The MITRE Corporation. For further information, please contact The MITRE Corporation, Contract Office, 7515 Colshire Drive, McLean, VA 22102 (703) 983-6000.

©2010 The MITRE Corporation. The Government retains a nonexclusive, royalty-free right to publish or reproduce this document, or to allow others to do so, for "Government Purposes Only".

Optimal Route Generation With Geometric Recourse Model Under Weather Uncertainty

Yoonjin Yoon

Ph.D. candidate

Dept of Civil and Environmental
Engineering, Univ. of California
Berkeley, CA, USA
yoonjin@berkeley.edu

Mark Hansen

Professor

Dept of Civil and Environmental
Engineering, Univ. of California
Berkeley, CA, USA

Michael O.Ball

Professor

Robert H. Smith School of Business,
Univ. of Maryland
College Park, MD, USA

Abstract— There has been growing interest in air transportation community to develop a routing decision model based on probabilistic severe weather. In the probabilistic air traffic management (PATM), decisions are made based on the stochastic weather information in the expected total cost sense. In this paper, we propose a geometric model to generate optimal route choice to hedge against weather risk. The geometric recourse model (GRM) is a strategic PATM model that incorporates route hedging and en-route recourse to respond to weather change. Hedged routes are routes other than nominal or detour route, and aircraft is re-routed to fly direct to the destination, or *recourse*, when the weather restricted airspace become flyable. Aircraft takes either the first recourse or the second recourse. The first recourse occurs when weather clears before aircraft reaches it when flying on the initial route. The second recourse occurs when the aircraft is at the weather region. There are two variations of GRM: Single Recourse Model (SRM) with first recourse only and Dual Recourse Model (DRM) with both the first and second recourse. When the weather clearance time follows a uniform distribution, SRM becomes convex with optimal solution is either at the upper bound or interior. Convexity gives optimality conditions in a closed form and analytic interior solution is approximated with marginal error. We prove that DRM has an important property such that when the maximum storm duration time is less than the flight time to the tip of the storm on detour route, it is always optimal to take the nominal route. Numerical study shows a substantial cost saving from using geometric recourse model, especially with DRM. It also indicates the need to consider ground holding in combination of route hedging.

Keywords-ATM; PATM; stochastic optimization; geometric model; risk hedging; severe weather event

I. INTRODUCTION

There is a growing interest in air traffic management (ATM) strategies that incorporate uncertainty in the national airspace system (NAS). Research in “probabilistic air traffic management” (PATM) seeks to guide decisions on ground-holding or otherwise modifying aircraft four-dimensional trajectories (4DTs) in order to minimize the expected cost, or to hedge against “worst case” scenarios in the next generation air transportation system.

This research studies the problem of developing a minimum-cost aircraft routing strategy when some weather condition inhibits the use of nominal route for an indefinite

period. In conventional Air Traffic Management (ATM), two options are commonly considered in this situation; the flight is either held at the origin airports until the nominal route becomes flyable or rerouted to avoid the weather region. The choice between these options is based upon a deterministic and conservative characterization of future weather, often resulting in underutilized airspace and unnecessary delay if the weather clears early.

In this paper, we propose a geometric model to find an optimal route when the weather clearance time is stochastic. The route decision takes into account the probability distribution of storm clearance times, the possibility of route hedging, and recourse opportunities. When facing uncertain weather, there are two potential risks to hedge against: persistence risk and clearance risk. Persistence risk is the risk when we take an “optimistic” route and weather persists, resulting in unplanned re-routing and delay. Clearance risk is the risk when we take a “pessimistic” route and weather clears sooner, resulting in unnecessary flight time. To mitigate these risks, we need to consider intermediate routing options that may not be chosen under either persistence or clearance, but hedge against either possibility. In addition, we must consider how the route might be adjusted if the storm clears during the course of flight. We assume that the flight plan can be amended in such event so that the plane can go direct to the destination.

In our model, the routing decision is made based on four parameters; nominal route between origin and destination airport, storm location, storm size, and maximum storm duration time. The optimistic route is the nominal one while pessimistic route goes around the storm. A hedged route is one that is between the optimistic and the pessimistic ones. We use the term “recourse” for a change in a routing that results from the storm clearing. We consider two recourse possibilities. First, the storm may clear before the aircraft reaches it, so that it can be rerouted directly to its destination. This is called the first recourse. The storm may instead persist beyond the time when the aircraft reaches it—so that the plane must turn and fly around it, but clear before the tip of the storm is reached. The aircraft may then be rerouted direct to the destination; we refer to this as second recourse.

In our model, which we term the geometric recourse model (GRM), a triangle is drawn where the base is the nominal route between the origin and destination airport, and the vertex is the

tip of the storm, which we assume to be a straight line perpendicular to the nominal route. We seek routes that minimize expected total flight cost, which in some cases are hedged routes. We consider two variations of geometric recourse model: the single recourse model (SRM) and dual recourse model (DRM). The SRM allows first recourse only, while the DRM allows both first and second recourse. Both SRM and DRM involve reroutes away from the weather region, while DRM includes reroutes in that region as well. The SRM is more conservative, while the DRM is more flexible and results in additional cost saving.

This paper introduces the concept of geometric recourse model and formulates nonlinear stochastic optimizations for the SRM and DRM. We assume that the storm clearance time follows a uniform distribution. With this assumption, we show that the SRM becomes convex, and find optimality conditions and the approximate analytic solution in closed form. We also find the condition that guarantees the nominal route to be optimal in DRM. Through numerical study, we compare the total expected flight cost and cost saving for optimal routes obtained from the SRM and DRM under a wide range of parameter values.

II. BACKGROUND

While traffic in the national airspace system has temporarily abated, its pre-recession level was approaching the capacity limit, with air travelers frequently experiencing flight delays and cancellations. Out of all causes of such delays, weather has been the most dominant one. According to the US Department of Transportation, air travelers experienced the worst flight delay in 2007 since year 2000, and weather accounted for more than 75% of these delays, as shown in Fig 1.

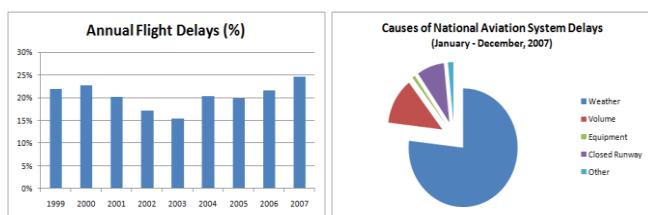


Figure 1. Annual Flight Delay Trend and Causes of System Delays in 2007
(Data Source: Bureau of Transportation Statistics)

In the event of adverse weather, one of the most widely used delay mitigation processes is the ground delay program (GDP). In a GDP, flights are held on the ground at the origin airport and assigned to new departure times based on available capacity at the destination airport. Although serving the purpose of handling the arrival capacity restrictions well, GDP are less well-suited for airspace capacity restrictions. Consequently, the Federal Aviation Administration (FAA) implemented the Airspace Flow Program (AFP) in June 2006. The purpose of AFP is to control the en-route traffic demand in regions of airspace that are capacity-constrained, most commonly as the result of severe weather.

Neither GDPs nor AFPs explicitly recognize that future weather is uncertain. As a result, when weather changes unexpectedly, a significant amount of reactive and tactical control is required, often resulting in inefficient system utilization. The motivation of this research is to integrate probabilistic weather information into strategic planning to provide flexible and effective decision support in order to reduce losses from imperfect information about future weather.

III. LITERATURE REVIEWS

There have been numerous efforts to address weather-related disruptions in the air traffic management. Earlier traffic flow management models such as Bertsimas [1] and Goodhart [2], often have a deterministic setting.

More recently, Nilim et al. [4] proposed a dynamic aircraft routing model with robust control. This paper adopted shortest-path algorithms in a grid structure, by discretizing time into stages when the routing decisions are made, and airspace as a two-dimensional grid. The weather condition in each potential storm region is assumed and modeled as a Markovian process with two states: 0 (No storm) and 1 (Storm). The transition matrix is estimated based on the historical weather forecasts. Optimization results show a promising improvement compared to flying around the storm without recourse.

The method in their paper has a robust control algorithm that has a wide range of applications. In the air transportation system however, the frequent routing adjustments entailed by this approach may place undue workload on controllers and pilots. Moreover, the Markovian assumption is of doubtful validity in the context of convective weather. Two of the goals in our study are to set up a model that has the flexibility to adopt a variety of probability distributions of storm clearance times, and to limit re-routing decision to a reasonable number.

Bertsimas et al. [3] proposed a two-stage optimization model based on a dynamic network flow approach. The authors set up a multi-aircraft optimization model minimizing the weather delay cost, based on a deterministic weather scenario. One important aspect of their study is that the cost function covers all phases of aircraft operation costs, such as fixed cost, ground holding cost, aircraft availability, and airborne cost.

From the air traffic management perspective, it would be ideal to utilize both Ground Delay Program (GDP) and airborne rerouting to mitigate weather related disruptions, especially since ground delay is less costly than extra flight time. Here, we do not explicitly consider the ground delay option, but instead focus on the choice of routing for a given time of departure. The extension of the model to support choice among alternate departure times is discussed at the conclusion of this paper.

IV. GEOMETRIC REOURSE MODEL (GRM)

A. Geometric Recourse Model Concept

Consider the problem of routing a single flight in the presence of a single storm. Given the origin and destination pair, assume there is a linear storm of known size blocking the direct route at a certain location. Using these five parameters-origin (O), destination (D), storm-route intersection (S_L), and

storm tip (S_T), construct a triangle ODS_T , where the nominal route is the base \overline{OD} and storm size is the altitude $\overline{S_L S_T}$, as illustrated in Fig 2. Note that while the storm has two tips, we choose the one nearer to SL , since this is the one that the aircraft would be routed around. Defining the unit of distance such that the aircraft cruises at a constant speed of 1, we refer to the base \overline{OD} as the nominal route, the altitude $\overline{S_L S_T}$ as the front of the storm and the vertex ST as the tip of the storm. The route $\overline{OS_T D}$, which goes around the storm, is called the *detour* route. Upon departure, the aircraft may set a course along the nominal route, the detour route, or one in between.

During the course of the flight, aircraft may be re-routed to fly direct to the destination when the storm clears; we refer to such route changes as *recourse*. Depending on the timing of storm clearance, there are three recourse possibilities as illustrated in Fig 3: (a) recourse if the storm clears before the aircraft reaches it; (b) recourse at the storm front if the storm persists until after the aircraft reaches it, but clears as the aircraft flies along the storm front toward the tip; or (c) no recourse because the storm persists until after the aircraft reaches the tip of the storm. We define the case (a) as the *first recourse*, the case (b) as the *second recourse*, and the case (c) as *no recourse*. Given the geometric setup, the objective is to find the route that minimizes expected total flight cost. Choosing a route is equivalent to choosing an angle between zero and the base angle $\angle S_T OS_L$. Although such a decision variable is intuitive, the resulting objective function involves complex trigonometric terms that make it difficult to analyze. Instead, we propose a ratio-based model in which complexity is reduced without loss of generality.

In the ratio-based model, the nominal route and weather parameters are expressed as ratios to the nominal route as illustrated in Fig 4. In other words, we define the unit of distance as the length of the nominal route, and the unit of time as the time required to fly that route. We also introduce a new decision variable x , which is the distance from the origin to the storm front along the course set from the origin. The ratio-based model is then formulated as follows.

- 1: length of nominal route between origin and destination
- β : storm size in units for nominal route length
- α : ratio of storm distance from origin in units of nominal route length: $0 < \alpha < 1$
- μ : random variable representing the storm clearance time with probability density function $p(\mu)$
- x : distance to the storm along course set from origin in units of nominal route length: $\alpha \leq x \leq \sqrt{\alpha^2 + \beta^2}$.

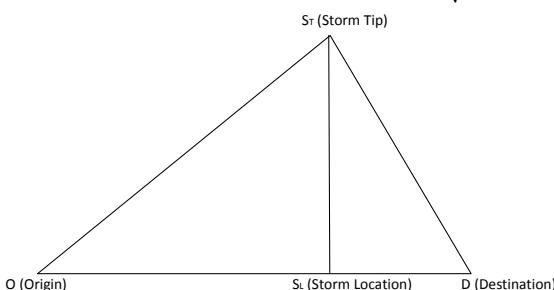


Figure 2. Geometric Model Concept

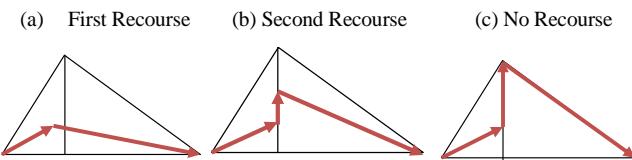


Figure 3. First, Second and No Recourse

- (a) *First recourse*: Recourse before the storm front if the storm clears before the aircraft reaches the edge.
- (b) *Second recourse*: Recourse in the storm region if the storm clears while the aircraft flies along the edge but before reaching the tip of the storm.
- (c) *No Recourse*: If storm doesn't clear until the aircraft reaches its tip, then fly around the storm.

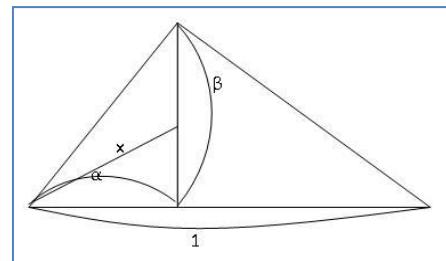


Figure 4. Ratio Based Geometric Model

As noted above, we consider two variations of geometric recourse model – Single Recourse (SRM) and Dual Recourse (DRM). The DRM, because it allows for immediate rerouting of flights moving along the storm region when the storm burns off, is more responsive. The SRM, because it assumes a large penalty for flights that reach the storm region prior to storm clearance, is more conservative. The SRM also has value from pure modeling point of view, since it provides an upper bound to DRM.

B. Single Recourse Model (SRM)

Single Recourse Model (SRM) is a geometric recourse model with first recourse only. The optimization model is formulated as follows.

Decision Variable: x

Objective Function:

$$\begin{aligned} \min & \int_0^x \left(\mu + \sqrt{1 + \mu^2 - 2\mu \frac{\alpha}{x}} \right) p(\mu) d\mu \\ & + \int_x^\infty \left(x \right. \\ & \left. + \sqrt{(x + \beta - \sqrt{x^2 - \alpha^2})^2 + \sqrt{(1 - \alpha)^2 + \beta^2}} \right) p(\mu) d\mu \\ \text{s.t. } & \alpha \leq x \leq \sqrt{\alpha^2 + \beta^2} \text{ where } 0 < \alpha < 1, \beta > 0. \end{aligned}$$

In the objective function, the first integral represents the expected total flight cost when first recourse is taken, and the second integral is the case when no recourse is possible. In the following section, we prove that SRM becomes convex and identify optimality conditions as well as approximated analytic solution when weather follows a uniform distribution.

C. Dual Recourse Model (DRM)

Dual recourse model allows recourse both before and at the storm region, providing most flexible environment. The optimization model formulation is as follows:

Decision Variable: x

Objective Function:

$$\begin{aligned} \min & \int_0^x \left(\mu + \sqrt{1 + \mu^2 - 2\mu \frac{\alpha}{x}} \right) p(\mu) d\mu \\ & + \int_x^{x+\beta-\sqrt{x^2-\alpha^2}} \left(\mu \right. \\ & \left. + \sqrt{\left(\mu - x + \sqrt{x^2 - \alpha^2} \right)^2 + (1 - \alpha)^2} \right) p(\mu) d\mu \\ & + \int_{x+\beta-\sqrt{x^2-\alpha^2}}^{\infty} \left(x \right. \\ & \left. + \sqrt{\left(x + \beta - \sqrt{x^2 - \alpha^2} \right)^2 + \sqrt{(1 - \alpha)^2 + \beta^2}} \right) p(\mu) d\mu \end{aligned}$$

$$\text{s.t. } \alpha \leq x \leq \sqrt{\alpha^2 + \beta^2} \text{ where } 0 < \alpha < 1, \beta > 0.$$

In the objective function, the first integral is the expected total flight cost when first recourse is taken, the second integral is the case when the second recourse is taken, and the third integral is the case when no recourse was possible.

We introduced the concept and two variations of geometric recourse model (GRM). In the next section, we present analytic study to find optimality conditions and an approximation of optimal solution in a closed form. A discussion on several important properties of GRM is followed. Performances of two models are compared and discussed in the following numerical analysis section.

V. ANALYTIC STUDY

A. Uniform Weather Distribution

We assume that the weather (storm) clearance time follows a uniform distribution ranging between 0 and T , or $\mu \sim \text{Uniform} [0, T]$. Forecast on convective activities in the airspace is included in several weather forecast products published by National Oceanic and Atmospheric Administration (NOAA)'s Storm Prediction Center (SPC). One of the widely used forecasts both in practice and in research is the convective outlook watch. According to SPC, they publish roughly 1,000 watches each year to address possible severe weather condition in the next few hours, and each convective activity is associated with a probability. Uniform distribution can utilize the single probability provided in the forecast, and be easily updated with new information as a new watch or warning is published.

With the uniform distribution assumption, we now have an additional parameter T , which is the latest possible time that storm will remain, or maximum storm duration time. Note that in the ratio-based optimization model, T is the actual maximum storm duration time divided by nominal route flight time. With the introduction of T , it is clear that we have a trivial solution $x^* = \alpha$, if $T \leq \alpha$.

B. Single Recourse Model (SRM)

1) Convex Optimization and Optimality Conditions

With the uniform weather distribution assumption, SRM becomes convex with negative gradient at the lower bound. We confirm the convexity by showing that the minimum of the second derivative of the objective function is positive. Let $f_s(x)$ be the expected total cost of SRM given weather parameters α, β and T . We first obtain the second derivative of $f_s(x)$ with respect to x and call it f_{sx} . To find the minimum of the second derivative with all possible weather parameter values, we treat f_{sx} as a function of x as well as α, β, T as shown below.

$$\begin{aligned} \min & \frac{\partial^2}{\partial x^2} f_{sx}(x, \alpha, \beta, T) \\ \text{s.t. } & \alpha \leq x \leq \sqrt{\alpha^2 + \beta^2}, 0 < \alpha < 1, \beta > 0, T > 0 \end{aligned}$$

We can solve the optimization model numerically which gives the global optimum of zero.¹

Now we show that the gradient at the lower bound is negative as follows.

$$f'_s(\alpha) = \infty \frac{\text{sign}((-1+\alpha)(-1+\sqrt{(-1+\alpha)^2+\alpha}))}{\text{sign}(T)} < 0. \quad (1)$$

Now, the SRM optimality condition is summarized as follows.

$$x^* = \begin{cases} \in (\alpha, \sqrt{\alpha^2 + \beta^2}), & f'_s(\sqrt{\alpha^2 + \beta^2}) > 0 \\ \sqrt{\alpha^2 + \beta^2}, & f'_s(\sqrt{\alpha^2 + \beta^2}) \leq 0 \end{cases} \quad (2)$$

In other words, if the gradient at the upper bound is positive, then there exists an interior solution. Otherwise, the upper bound of x is the optimal solution. Interior solution is equivalent to taking a route inside the triangle and represents the case when hedging is optimal.

Rearranging the first condition in (2) yields that $f'_s(\sqrt{\alpha^2 + \beta^2}) > 0$ is equivalent to $T < \theta_s(\alpha, \beta)$.² The formula for $\theta_s(\alpha, \beta)$ is quite complex to interpret in its analytic form. Instead, the condition $T < \theta_s(\alpha, \beta)$ is represented in a contour map in the α - β plane in Fig 5. In the contour map, each contour line corresponds to a value of $\theta_s(\alpha, \beta)$. Using this map, one can determine whether there is an interior solution or not once the weather parameters are known. For example, if $\alpha=0.6$ and $\beta=0.4$, there is an interior solution when $T=1$, and no interior solution when $T=2$. Note that as T gets larger, it is less and less likely to have an interior solution, which matches our intuition. The contour map is a valuable decision reference to determine whether hedging is worth considering or not without solving optimization.

¹ The global optimum is unique at zero although there are multiple optimal solutions.

² See Appendix for complete formula of $\theta_s(\alpha, \beta)$

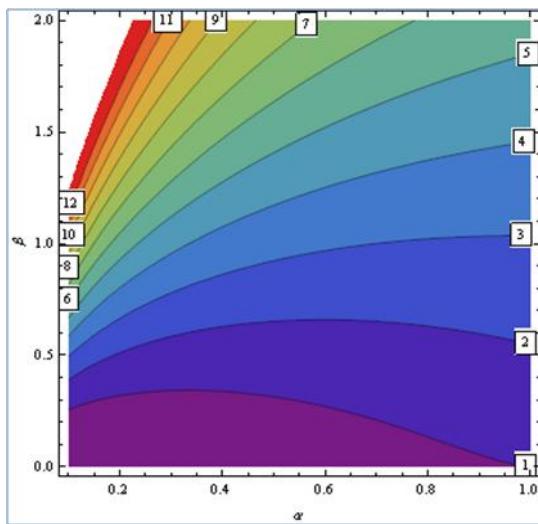


Figure 5. Condition of Interior Solution of SRM

Once the weather parameters are known, one can refer to this plot to decide whether it is worthwhile to find an optimal angle below the maximum angle, or it would be best to simply fly around the storm. Number in the white squares is the value of $\theta_s(\alpha, \beta)$. For example, with $\alpha=0.4$ and $\beta=0.5$, it is optimal to fly around the storm if $T=3$. On the other hand, if $T=2$, then there is an interior solution. Note that as T gets larger, it is less likely to have an interior solution.

2) Analytic Solution Approximation using Taylor Series

The optimization problem doesn't have an analytic solution. We apply Taylor series expansion to approximate our objective function as a polynomial of degree 2 in x , around the middle point of its domain $(\alpha + \sqrt{\alpha^2 + \beta^2})/2$.

The Taylor series approximation, which we call $f_{ts}(x)$ is quite complex, but we're only interested whether the minimizing x falls inside the domain or not. In other words, the interior solution is approximated as $-\frac{\text{Coefficient}(f_{ts}, 1)}{\text{Coefficient}(f_{ts}, 2)}$, where $\text{Coefficient}(f, n)$ denotes the coefficient of x^n of polynomial function f . The optimal solution is summarized in Eq. (3).

$$x^* = \begin{cases} \alpha & T \leq \alpha \\ -\frac{\text{Coefficient}(f_{ts}, 1)}{\text{Coefficient}(f_{ts}, 2)} & \alpha < T < \theta_s(\alpha, \beta) \\ \sqrt{\alpha^2 + \beta^2} & T \geq \theta_s(\alpha, \beta) \end{cases} \quad (3)$$

We tested our model with different weather parameters and approximation error was less than 1% in most cases. A sample results are shown in the Table I.

TABLE I. APPROXIMATION ERROR SAMPLES

	Geometry	E(Total Cost)	Int. sol	x^*	x^{*ts}	error
$\alpha=0.2$ $\beta=0.5$ $T=2.5$			N	0.53	0.53	0%
$\alpha=0.5$ $\beta=0.5$ $T=1.1$			Y	0.66	0.64	2%
$\alpha=0.8$ $\beta=0.5$ $T=1.1$			Y	0.85	0.85	0.3%

C. Dual Recourse Model (DRM)

Dual Recourse Model (DRM) is neither always convex nor concave and its properties are best addressed in our numerical analysis in the following chapter. However, DRM has an important property such that when the maximum storm duration time is relatively short, taking nominal route is always optimal.

Theorem. In DRM, $x^* = \alpha$ if $0 < T \leq \sqrt{\alpha^2 + \beta^2}$.

Proof. It is trivial that $x^* = \alpha$, when $T \leq \alpha$. (4)

If $\alpha < T \leq \sqrt{\alpha^2 + \beta^2}$, the objective function $f_d(x)$ is as follows.

$$f_d(x) = \int_0^x \left(\mu + \sqrt{\mu^2 + 1 - 2\mu \frac{\alpha}{x}} \right) p(\mu) d\mu + \int_x^T \left(\mu + 1 - \alpha^2 + \mu - x + x^2 - \alpha^2 \right) p\mu d\mu \quad (5)$$

Then,

$$\begin{aligned} f_d(x) - f_d(\alpha) = & \int_0^\alpha \left(\sqrt{\mu^2 + 1 - 2\mu \frac{\alpha}{x}} - \sqrt{\mu^2 + 1 - 2\mu} \right) p(\mu) d\mu + \\ & \int_\alpha^x \left(\sqrt{\mu^2 + 1 - 2\mu \frac{\alpha}{x}} - 1 - \alpha^2 + \mu - \alpha^2 \right) p\mu d\mu - \\ & - 1 - \alpha^2 + \mu - \alpha^2 p\mu d\mu + xT1 - \alpha^2 + \mu - x + x^2 - \alpha^2 \end{aligned} \quad (6)$$

To show $f_d(x) - f_d(\alpha) > 0, \forall x \in (\alpha, T]$, we show that each integrand in (6) is non-negative. It is trivial that

$$\sqrt{\mu^2 + 1 - 2\mu \frac{\alpha}{x}} - \sqrt{\mu^2 + 1 - 2\mu} \geq 0. \quad (7)$$

Since $\left(\mu^2 + 1 - 2\mu \frac{\alpha}{x} \right) - ((1 - \alpha)^2 + (\mu - \alpha)^2) > 0$ when $\alpha < \mu \leq x$, we have

$$\sqrt{\mu^2 + 1 - 2\mu \frac{\alpha}{x}} - \sqrt{(1 - \alpha)^2 + (\mu - \alpha)^2} > 0. \quad (8)$$

Similarly,

$$\sqrt{(1 - \alpha)^2 + (\mu - x + \sqrt{x^2 - \alpha^2})^2} - \sqrt{(1 - \alpha)^2 + (\mu - \alpha)^2} > 0 \quad (9)$$

, where $x < \mu \leq T$.

From (7), (8) and (9), we have $f_d(x) - f_d(\alpha) > 0, \forall x \in (\alpha, T]$. Therefore,

$$x^* = \alpha \text{ if } 0 < T \leq \sqrt{\alpha^2 + \beta^2}. \quad (Q.E.D.)$$

This theorem identifies the condition to always choose the nominal route regardless of the weather probability distribution, when the storm is expected to last relatively for a short amount of time. In other words, if the maximum storm duration time is less than the time to fly to the tip of the storm on detour route, then it is always optimal to fly on the nominal

route. This theorem provides the condition of critical cost saving opportunity without even considering route hedging.

So far, we studied formulation and properties of the geometric recourse models. Another primary interest is to measure the improvement from adopting DRM or SRM. Since DRM is an upper bound to SRM, DRM guarantees less cost than SRM. Likewise SRM guarantees less cost than simply taking the detour route. In the next section, we discuss various performance metrics based on numerical analysis.

VI. NUMERICAL STUDY

In numerical study, a 3-D grid structure is created in α - β - T plane with each grid being a cube with side of 0.05, or 5% of the nominal route. We also set reasonable limit to storm size and maximum storm duration time as 2 and 4 respectively. Therefore, $\alpha \in [0.05, 0.95]$, $\beta \in [0.05, 2)$, $T \in [0.05, 3]$. Note that when $\alpha=0$ or $\alpha=1$, the storm is located at the origin or the destination airport, in which case ground delay program works best.

In numerical analysis, we consider three scenarios as detailed below. The available options for each scenario are summarized in Table II.

- *Baseline*: take detour route and whenever storm clears before aircraft reaches the storm tip, fly direct to the destination
- *SRM*: utilize SRM with first recourse option only
- *DRM*: utilize DRM with both first and second recourse options.

It is clear that the baseline case is an upper bound of SRM, which is then an upper bound of DRM. For each (α, β, T) , we find solutions for these three cases and obtain performance metrics.

In the following section, we compare optimal cost and cost saving of those three cases. Note that if $T \leq \alpha$, all three cases yield the same solution $x^* = \alpha$, and we exclude these trivial cases from our analysis.

TABLE II. OPERATION OPTIONS FOR BASELINE, SRM AND DRM

Model	Hedging	First Recourse	Second Recourse
Baseline	X	O	X
SRM	O	O	X
DRM	O	O	O

A. Minimum Expected Total Cost (ETC^*)

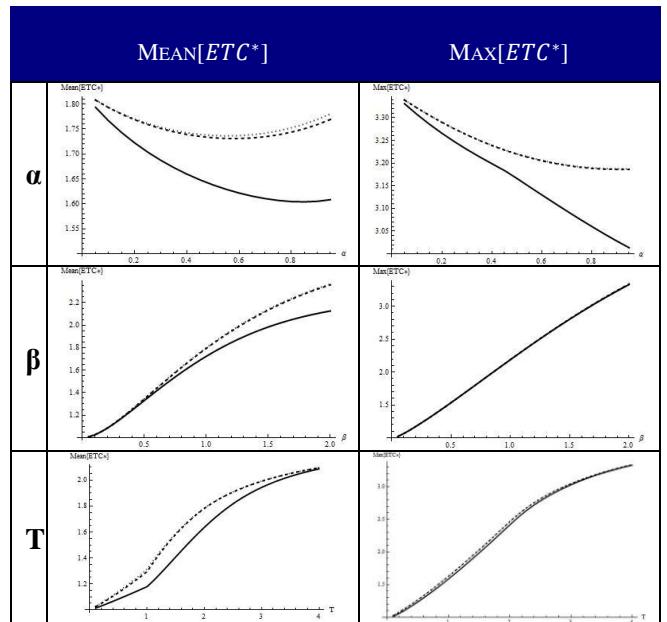
To study the optimal cost (ETC^*) with respect to each weather parameter α , β and T , we find the minimum, average and maximum of all ETC^* 's, when one of the parameters is fixed at a certain value. For example, to analyze ETC^* with respect to α , first set $\alpha=0.05$ and collect all ETC^* 's of baseline, SRM and DRM respectively and find minimum, average and

maximum of ETC^* 's for three cases. Continue for $\alpha=0.1, 0.15, \dots 0.95$. We repeat the same process for β and T , and the result is summarized in Table III.

$\alpha - Mean[ETC^*]$ plot shows how the average optimal cost changes with respect to α . We can see that on average, ETC^* of baseline and SRM is nearly the same and tends to reach its minimum when α is near 0.6. In other words, there is little difference in the average performance between baseline and SRM, while both of them performs best when α is around 0.6. As expected, DRM always performs better than baseline and SRM, and ETC^* gradually decreases as α increase until α is almost 1. We also observe for storm located near the destination, DRM performs much better than SRM or baseline. $\alpha - Max[ETC^*]$ plot shows the worst-case performances. We can see that all three cases show little difference when storm is very near the origin and optimal cost gradually decreases as storm moves toward the destination, especially for DRM.

In the β plots, we find that ETC^* increases as β increases, while average performance of DRM is better than the baseline or SRM. It matches our intuition since as the storm gets larger, second recourse option in DRM will pay off, although in the worst case when no recourse is possible, all three models will perform the same.

TABLE III. MEAN AND MAXIMUM OF ETC^* WITH RESPECT TO WEATHER PARAMETERS



Note: Baseline case is shown in dotted line, SRM is shown in dashed line, and DRM is shown in solid line.

In summary, there is little difference between SRM and baseline case, while DRM sometimes performs substantially better. On average, expected total cost increases with decreasing rate as the storm size and the maximum duration time increases. On the other hand, expected total cost is convex with respect to the storm location, as it decreases up to a minimum point then increases. We observe that DRM reduces weather risk further with a storm located near the destination airport. This is also true when storm size is large. There is a

range of T where DRM achieves substantially less expected total cost, although such advantage disappears as T becomes very large.

B. Cost Saving

Let's define cost saving of SRM and DRM as follows.

$$S(SRM) = 1 - \frac{ETC^*[SRM]}{ETC^*[Baseline]} \quad (4)$$

$$S(DRM) = 1 - \frac{ETC^*[DRM]}{ETC^*[SRM]} \quad (5)$$

, where $ETC^*[\cdot]$ is the minimum expected total cost of the selected model.

The cumulative distribution functions of $S(SRM)$ and $S(DRM)$ are shown in Fig. 6. With SRM, nearly 90% of cases have less than 1% saving compared to the baseline case, and more than 99% cases has less than 5% saving with the largest saving close to 6%. With DRM, more than 32% has savings larger than 5% with largest saving close to 30%. In fact, about 20% shows significant saving larger than 10%.

In Table IV, the average, minimum and maximum cost savings are plotted with respect to each parameter. Maximum and average savings of both α -S(SRM) and α -S(DRM) are monotonic increasing functions. The convex shape of α -S(SRM) plot suggests that SRM works best with storms very near the destination, while the concave shape of α -S(DRM) plot suggests that DRM works well with wide range of storms as well as those very near the destination. We also observe the average cost saving of SRM is negligible.

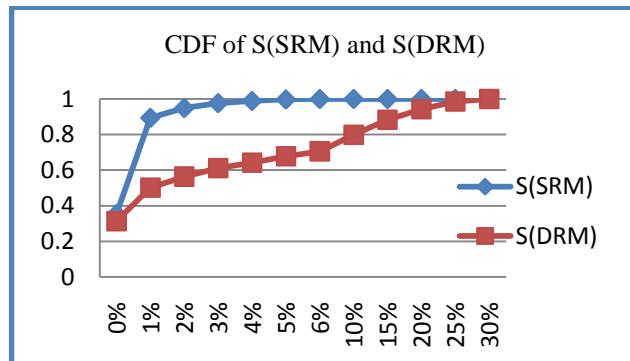


Figure 6. . Cumulative Distribution Function of Cost Saving of SRM (S(SRM)) and Cost Saving of DRM (S(DRM))

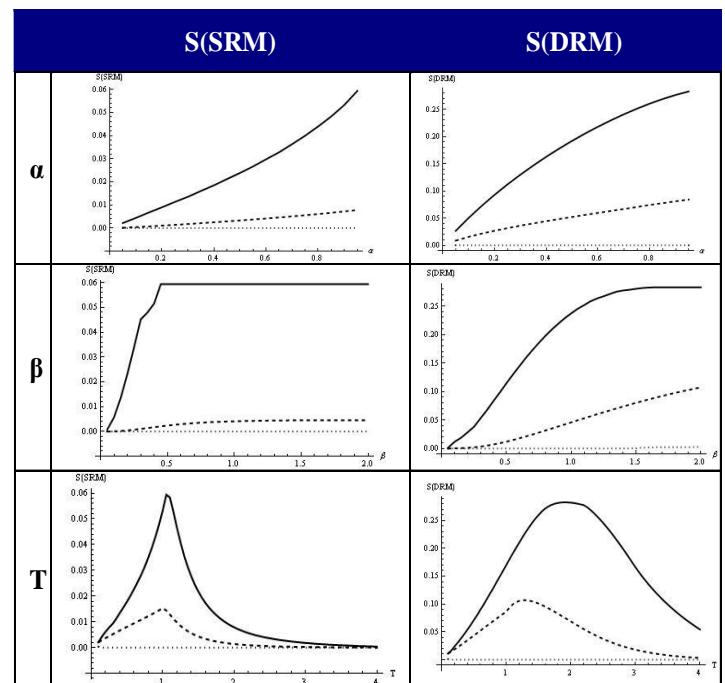
β -S(SRM) plot shows diminishing returns, cost saving increases until β reaches near 0.5 then stays flat afterward. It is intuitive that without second recourse option, SRM is limited to hedge the risk of larger storm. Compared to β -S(SRM), β -S(DRM) plot shows wide range of β that DRM saves meaningful cost, which coincides with our finding in the previous section. More importantly, the average cost saving maintains increasing trend even after maximum saving plateaus, which suggests that DRM is effective in reducing the risk of large-sized storms.

In both T -S(SRM) and T -S(DRM) plots, there are ranges of T showing the peak savings. The largest saving of SRM is close to 6% when T is between 1 and 1.2, and it is close to 30% when T is between 1.9 and 2.1 for DRM. There is an

interesting observation when it comes to average cost saving. For SRM, maximum average cost saving of T -S(SRM) is much higher than those of α -S(SRM) and β -S(SRM).

Although it appears that the average cost saving of SRM is negligible in α -S(SRM) plot, there are cases when it becomes meaningful when T is in a certain range. We make similar observation for DRM as well. Such observations suggest that performance of these geometric recourse models is more dependent to the maximum storm duration than the location or the size of storm. It also suggests to consider ground delay in combination with route hedging, which essentially reduces the maximum storm duration time. Another important observation is the large gap between the average and maximum cost saving in general, which indicates that there exist certain combinations of α , β and T that these models show true advantage.

TABLE IV. MEAN AND MAXIMUM OF $S(SRM)$ AND $S(DRM)$ WITH RESPECT TO WEATHER PARAMETERS



Note: Minimum is shown in dotted line, average is shown in dashed line, and the maximum is shown in solid line.

VII. CONCLUSIONS AND FUTURE RESEARCH

In this paper, we propose a geometric recourse model to generate optimal route to hedge against weather risk in the airspace. The Geometric Recourse Model (GRM) incorporates weather risks into strategic planning and provides risk-hedging opportunity with intermediate routes as well as the nominal and detour route. It also provides added flexibility by rerouting aircraft to fly direct to destination, or *recourse*, as soon as the weather restricted airspace becomes flyable again. There are two recourse options we consider. The first recourse is the case when the weather clears before aircraft reaches its front, and the second recourse option is the case when weather persists when aircraft reaches it then clears while the aircraft is still in the weather region.

We consider two geometric recourse models; Single Recourse Model (SRM) with the first recourse option only and Dual Recourse Model (DRM) with both the first and the second recourse option. In SRM, the optimization model becomes convex when the weather probability follows a uniform distribution, and optimal solution occurs either in the interior or at the upper bound. Convexity gives optimality conditions in closed form. The interior solution is approximated with Taylor series with marginal error. SRM has significance from pure modeling point of view as well, since it provides an upper bound to DRM.

DRM has an important property that it is always optimal take the nominal route when the maximum storm duration time is less than the time to fly to the tip of the storm.

In numerical study, a 3D grid structure is created in α - β -T space, where α is the location, β is the size and T is the maximum duration time of the storm, and both models are solved for each tuple in the grid. Below is the summary of key findings from the numerical analysis.

- SRM works best with storms very near the destination and relatively in small size.
- DRM works well with wide range of storm location and larger storms.
- Cost saving distributions show that nearly 90% of cases we tested have less than 1% saving with SRM with largest possible saving close to 6%. On the other hand, almost 30% of all cases have larger than 10% saving with DRM with the largest saving reaching 30%.
- Both models show peak cost savings for T in a certain range. The maximum average cost saving is also higher for those T values, compared to the maximum average saving with respect to α and β . These observations suggest that the performance of our models is more sensitive to the maximum storm duration time than other two parameters, which gives us a strong motivation to consider ground delay in combination with route hedging, especially with storms expected to last longer.

As an immediate follow-up study, we currently study the value of hedging with various probability distributions. We

also plan to study ground-airborne hybrid model, where ground delay is another decision factor in addition to route choice.

ACKNOWLEDGMENT

This project is sponsored by NASA Ames Research Center, Mountain View, CA, USA.

REFERENCES

- [1] D. Bertsimas, S. Patterson, "The air traffic flow management problem with enroute capacity", *Operations Research*, 46:406-422,1998.
- [2] J. Goodhart, Ph. D thesis, Department of Industrial Engineering and Operations Research, University of California, Berkeley, 1999.
- [3] D. Bertsimas, S. Patterson, "The traffic flow management rerouting problem in air traffic control: a dynamic network flow approach", *Transportation Science*, INFORMS, Vol. 34, No. 3, 239-255, 2000.
- [4] S. Patterson, "The Traffic Flow Management Rerouting Problem in Air Traffic Control: A Dynamic Network Flow Approach", *Transportation Sciecne*, INFORMS, Vol. 34, No. 3,
- [5] A. Nilim, L. El Ghaoui, M. Hansen, V. Duong, "Trajectory-based Air Traffic Management under Weather Uncertainty", USA/EUROPE ATM R&D Seminars, 2001.
- [6] A. Nilim, L.El Ghaoui, V. Duong, "Robust dynamic routing of aircraft under uncertainty", Digital Avionics Systems Conference, 1A5-1-1A5-13 vol.1, 2002.
- [7] A. Nilim, L. El Ghaoui, "Algorithms for air traffic flow management under stochastic environments", American Control Conference, 3429-3434, 2004.
- [8] M. Ball, R. Hoffman, D. Lovell, "Response Mechanisms For Dynamic Air Traffic Flow", 6th USA/Europe Air Traffic Management R&D Seminar, 2005
- [9] "NBAA Navigator", Spring 2006
- [10] "Airspace Flow Program Frequently Asked Questions", Federal Aviation Administration, 2006.
- [11] "Ground Delay Program, Airspace Flow Program, And Substitution Message Processing V3.1", Federal Aviation Administration, 2008.
- [12] Weather Integrated Product Team, "Next Generation Air Transportation System: Weather Concept of Operations Ver 1.0", Joint Planning and Development Office
- [13] NOAA's Nation Weather Service Storm Prediction Center: <http://www.spc.noaa.gov/>
- [14] S. Boyd, "Convex Optimization, Course Reader", Department of Electrical Engineering and Computer Science, Stanford University.
- [15] S. Ross, "Stochastic Processes, Second Edition", John Wiley & Sons, Inc.
- [16] Folland, "Real Analysis, 2nd edition", Prentice Hall

Appendix

$$\begin{aligned}
\theta_s(\alpha, \beta) = & ((\alpha^2 + \beta^2)^2 (\alpha^5 + \alpha^4(-2 + \beta) + \beta^3 + \beta^5 + \alpha^3(1 - 3\beta + 2\beta^2) + \alpha^2\beta(3 - 2\beta + 2\beta^2) \\
& + \alpha\beta(-1 + \beta - 3\beta^2 + \beta^3)) \\
& - \sqrt{\alpha^2 + \beta^2} (\alpha^8 + \alpha^7(-2 + \beta) + 3\alpha^5(-1 + \beta)^2\beta + 3\alpha^3(-1 + \beta)^2\beta^3 + \beta^6 + \beta^8 + \alpha^6(1 - 3\beta + 4\beta^2) \\
& + \alpha^2\beta^3(1 + 3\beta - 3\beta^2 + 4\beta^3) + \alpha^4\beta(-1 + 3\beta - 6\beta^2 + 6\beta^3) + \alpha\beta^3(-1 - 2\beta^3 + \beta^4)) \\
& - \sqrt{1 - 2\alpha + \alpha^2 + \beta^2} (\alpha^8 + \alpha^7(-1 + \beta) + 3\alpha^3(-1 + \beta)\beta^4 + \beta^8 + 2\alpha^6\beta(-1 + 2\beta) \\
& + 2\alpha^2\beta^5(-1 + 2\beta) + 2\alpha^4\beta^3(-2 + 3\beta) + \alpha\beta^5(-1 - \beta + \beta^2) + \alpha^5\beta(1 - 3\beta + 3\beta^2)) \\
& + \sqrt{-2\alpha^3 + \alpha^4 + \beta^2 - 2\alpha\beta^2 + \beta^4 + \alpha^2(1 + 2\beta^2)} (\alpha^7 + \alpha^6(-1 + \beta) + \beta^7 + \alpha^3\beta^3(-4 + 3\beta) \\
& + \alpha^5\beta(-2 + 3\beta) + \alpha^4\beta(1 - 2\beta + 3\beta^2) + \alpha^2\beta^3(1 - \beta + 3\beta^2) + \alpha\beta^3(-1 - 2\beta^2 + \beta^3)) \\
& + \alpha^2\beta \left(\alpha^4 - \alpha^3 \left(2 + \sqrt{\alpha^2 + \beta^2} + \sqrt{1 - 2\alpha + \alpha^2 + \beta^2} \right) \right. \\
& \left. + \beta^2 \left(1 + \beta^2 + \sqrt{\alpha^2 - 2\alpha^3 + \alpha^4 + \beta^2 - 2\alpha\beta^2 + 2\alpha^2\beta^2 + \beta^4} \right) \right. \\
& \left. + \alpha^2 \left(1 + 2\beta^2 + 2\sqrt{\alpha^2 + \beta^2} + \sqrt{1 - 2\alpha + \alpha^2 + \beta^2} + \sqrt{\alpha^2 - 2\alpha^3 + \alpha^4 + \beta^2 - 2\alpha\beta^2 + 2\alpha^2\beta^2 + \beta^4} \right) \right. \\
& \left. - \alpha \left(\sqrt{\alpha^2 + \beta^2} + \sqrt{\alpha^2 - 2\alpha^3 + \alpha^4 + \beta^2 - 2\alpha\beta^2 + 2\alpha^2\beta^2 + \beta^4} \right. \right. \\
& \left. \left. + \beta^2 \left(2 + \sqrt{\alpha^2 + \beta^2} + \sqrt{1 - 2\alpha + \alpha^2 + \beta^2} \right) \right) \right) \left(\log \left[-\alpha + \sqrt{\alpha^2 + \beta^2} \right] \right. \\
& \left. - \log \left[-\alpha + \alpha^2 + \beta^2 + \sqrt{(\alpha^2 + \beta^2)(1 - 2\alpha + \alpha^2 + \beta^2)} \right] \right) / (\alpha^2 + \beta^2) \sqrt{1 - 2\alpha + \alpha^2 + \beta^2} (-\alpha + \alpha^2 \\
& + \beta^2 + \sqrt{(\alpha^2 + \beta^2)(1 - 2\alpha + \alpha^2 + \beta^2)}) (\alpha^3 + \alpha^2(\beta - \sqrt{\alpha^2 + \beta^2}) + \alpha\beta(\beta - \sqrt{\alpha^2 + \beta^2}) + \beta^2(\beta \\
& - \sqrt{\alpha^2 + \beta^2}))
\end{aligned}$$

A Diffusion Approximation to a Single Airport Queue

David Lovell and Kleoniki Vlachou
 University of Maryland
 CEE, ISR
 College Park, MD, USA
 [lovell, kvlachou]@umd.edu

Tarrek Rabbani and Alexandre Bayen
 University of California, Berkeley
 Civil and Environmental Engineering
 Berkeley, CA, USA
tarrek.rabbani@gmail.com, bayen@berkeley.edu

Abstract—This paper illustrates a continuum approximation to queuing problems at a single airport adapted from the well-known diffusion approximation, as encapsulated in the Kolmogorov forward equation of stochastic processes or the Fokker-Planck equation of physics. The continuum model is derived using special artifacts of the airport problem context. The appropriate initial and boundary conditions are defined and a numerical solution scheme based on the finite element method is presented.

Keywords—queuing theory; diffusion; delay; aviation system performance; Kolmogorov forward equation, Fokker-Planck equation; finite element method

I. INTRODUCTION

Studies of queuing delays in the National Airspace System (NAS), and other large networks, for that matter, are typically conducted either in a Monte Carlo simulation environment, where a considerable amount of fidelity is available at the expense of computational efficiency, or with closed-form equilibrium queuing models fraught with distributional assumptions that are typically not very representative of real situations. A common example of the latter is the use of the Poisson process to represent arrival processes to queues, motivated by its mathematical tractability, even in the face of fairly compelling evidence that the system is not Markovian.

With the aviation system in mind, the idea behind this paper is to adapt a somewhat common continuous approximation technique known as the diffusion approximation to a queuing problem, with a specific interest in modeling arrival and departure delay statistics at an airport over the course of several hours or a day. The primary advantages of using the diffusion approximation for these purposes are that specific distributional assumptions can be relaxed in favor of an approximate description of the relevant stochastic processes by a small number of their time-dependent moments, that the full spectrum of probabilistic results can be obtained via a single run of the model, and that propagation of higher moments beyond the mean queue behavior can be captured.

In general, we believe it should be possible to represent a network of queues using methodology similar to the methods herein, although the results to date apply only to a single queue with a general arrival and general service process. The presentation of the approach will continue as follows: in Section II, we show the derivation of the foundational partial differential equation that represents the system dynamics. That is followed by the development of the continuous equations

necessary to establish the boundary and initial conditions that assure the meaningful solution of a meaningful problem. In Section III, we show a numerical approximation scheme that is based on the finite element method (FEM), and is necessary to solve the problem by computer. In Section IV we close with some results illustrating the use of the model, and demonstrate some of the deleterious effects of numerical instability that can result from improperly scaled input data.

II. MODEL DEVELOPMENT

In this section we introduce the modeling assumptions that lead to the particular continuum approximation for queuing systems known as the diffusion approximation. This consists of a governing differential equation, which is presented first, which represents the primary dynamics of the system. This equation is valid for a closed subset of the real numbers representing all realistic values of the system state, but some boundary conditions must be imposed to prevent physically meaningless results outside of this interval. We also describe the set of initial conditions required to represent any particular queuing problem for which a solution is sought.

A. Governing Differential Equation

Diffusion methods have been applied to queuing problems in a variety of domains, including road transportation [1], computer networks [2], and more general queuing systems [3,4]. No significant use of them in an aviation setting is recorded in the literature. The development of the model shown in the following pages borrows very heavily from the exposition of Kimura [5], which develops the diffusion approximation in the context of a very different application, that of population genetics. The reason for following the template of that paper, however, is that the treatment is very thorough but also accessible to readers without prior experience in diffusion methods, and it can be adapted readily to the aviation context.

Suppose we model the arrival process to an airport as a single-server queue. Let $Q(t)$ represent the time-dependent random variable describing the length of the (virtual) queue for arrival aircraft at time t . Beyond the scope of this paper, the ultimate goal of this endeavor is to model more complicated aviation networks. As such, the airport node being described here might actually be an arrival or departure resource like a runway, it might be a gate, or it might be an esoteric en route node intended to represent a capacity constraint in the airspace itself.

The first assumption necessary for consideration of continuum models is that of continuity; i.e., that the queue length measurement at any given time need not be an integer. Because aircraft only come in discrete units, this is obviously an artificial construct. However, we are mostly interested in using queue length measurements as preliminaries to computing delay statistics, so they will be averaged over a large domain. As a result, this assumption is probably no more malignant than assuming that there is such a thing as a "queue" at an arrival airport. This is a stochastic queuing system, and the density function for the queue length at time t is denoted $f(x; t)$. A graphical example of f is shown in Fig. 1.

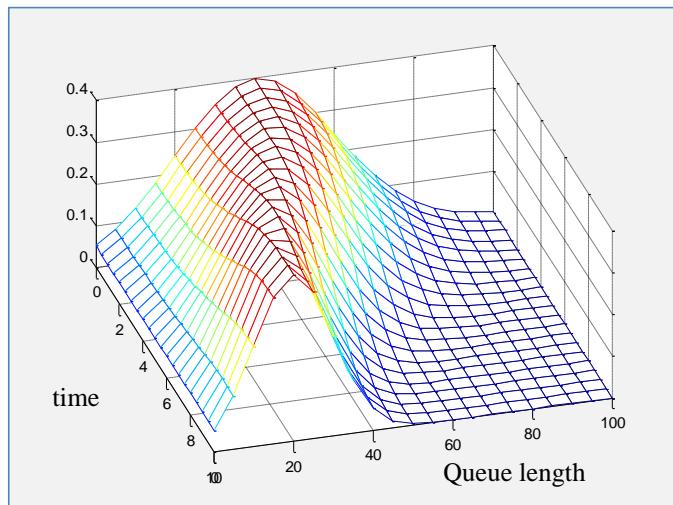


Fig. 1. Queue length probability density function

We also define the probability density transition function $g(\delta x, x; \delta t, t)$ as the probability density, associated with a change in queue length from x to $x + \delta x$ in the time interval $[t, t + \delta t]$. An example of g for a single choice of t and δt is shown in Fig. 2.

The density function for the queue length at some future time $t + \delta t$ can be expressed using the continuous version of the Kolmogorov-Chapman equation:

$$f(x; t + \delta t) = \int f(x - \delta x; t) g(\delta x, x - \delta x; \delta t, t) d(\delta x) \quad (1)$$

This equation encapsulates conditioning over all of the possible queue states at time t from which a transition to the state x at time $t + \delta t$ is possible. The necessary assumption is that the transition probabilities can be described entirely by the function g , regardless of the history of the prior queue states. Thus, the system can be described as Markovian.

If we use the condensed notation $fg = f(x; t)g(\delta x, x; \delta t, t)$, then we can expand the integrand of (1) as a Taylor series around the point x as follows:

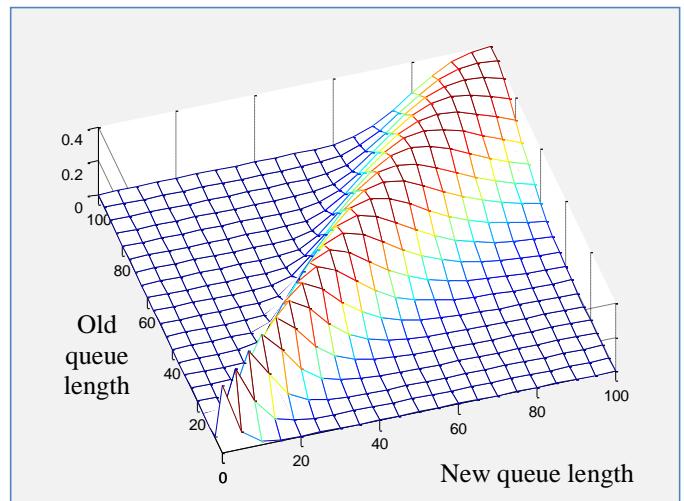


Fig. 2. State transition probability function

$$\begin{aligned} f(x - \delta x; t)g(\delta x, x - \delta x; \delta t, t) &= fg - \delta x \frac{\partial}{\partial x}(fg) + \\ &\frac{(\delta x)^2}{2!} \frac{\partial^2}{\partial x^2}(fg) - \frac{(\delta x)^3}{3!} \frac{\partial^3}{\partial x^3}(fg) + \dots \end{aligned} \quad (2)$$

We then substitute (2) back into (1), and interchange integration and differentiation. This presumes, of course that our functions are well-behaved in this sense.

$$\begin{aligned} f(x; t + \delta t) &= f \int g d(\delta x) - \frac{\partial}{\partial x} \left\{ f \int (\delta x) g d(\delta x) \right\} + \\ &\frac{1}{2} \frac{\partial^2}{\partial x^2} \left\{ f \int (\delta x)^2 g d(\delta x) \right\} - \dots \end{aligned} \quad (3)$$

Since g is a proper density function, then for any choices x , t , and δt , it must be the case that $\int g d(\delta x) = 1$. Hence we simplify the first term on the RHS of (3), and then subtract f from both sides and divide by δt :

$$\begin{aligned} \frac{f(x; t + \delta t) - f(x; t)}{\delta t} &= - \frac{\partial}{\partial x} \left\{ f(x; t) \frac{1}{\delta t} \int (\delta x) g d(\delta x) \right\} \\ &+ \frac{1}{2} \frac{\partial^2}{\partial x^2} \left\{ f(x; t) \frac{1}{\delta t} \int (\delta x)^2 g d(\delta x) \right\} - \dots \end{aligned} \quad (4)$$

The limits of two of the elements contained in the RHS of (4) are frequently called the "infinitesimal" mean and variance, respectively:

$$\lim_{\delta t \rightarrow 0} \frac{1}{\delta t} \int (\delta x) g(\delta x, x; \delta t, t) d(\delta x) \equiv M(x; t) \quad \forall x, t \quad (5)$$

$$\lim_{\delta t \rightarrow 0} \frac{1}{\delta t} \int (\delta x)^2 g(\delta x, x; \delta t, t) d(\delta x) \equiv V(x; t) \quad \forall x, t \quad (6)$$

The next assumption is that all of the important information about the transition density function g can be captured in its first and second moments, as in (5) and (6), respectively. This is not a severe limitation; for situations where this is not the case, additional infinitesimal moments can be defined, and the analyst is then responsible for providing that information as well. In fact, in aviation applications, the best contemporary network models only deal with the propagation of average behavior, so including $V(x;t)$ is already a step forward. For the present case, assuming that the first two moments suffice, this is tantamount to the assumption:

$$\lim_{\delta t \rightarrow 0} \frac{1}{\delta t} \int (\delta x)^n g(\delta x, x; \delta t, t) d(\delta x) = 0 \quad n \geq 3, \forall x, t \quad (7)$$

Then, taking the limit of (4) as $\delta t \rightarrow 0$ and substituting (5) and (6) yields:

$$\frac{\partial f(x;t)}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial x^2} V(x;t) f(x;t) - \frac{\partial}{\partial x} M(x;t) f(x;t) \quad (8)$$

Equation (8) is commonly called the *Kolmogorov forward equation* in the stochastic processes literature, or the *Fokker-Planck equation* in the physics literature. In the second case, the term $M(x;t)$ is referred to as *drift*, while the term $V(x;t)$ is called *diffusion*. Equation (8) is the governing differential equation (GDE) for our queuing system.

B. Boundary Conditions

In this section, we develop the boundary conditions that prevent the model from generating non-zero probabilities for states that are not physically possible, including negative values of the queue length. A similar constraint can be imposed to prevent the possibility of what might be considered unnaturally large queue lengths. It is much more difficult to specify this boundary precisely, but it is necessary from a pragmatic standpoint in the numerical scheme because the solution space must be bounded, as will be seen in Section III.

Because the random variable $Q(t)$ represents a queue length, it makes no sense for it to be negative. Thus, we want to establish an auxiliary condition that can be applied, in addition to (8), that guarantees that

$$f(x;t) = 0 \quad x < 0, \quad \forall t \quad (9)$$

This cannot be accomplished by simply saying that (9) must be true; an additional differential equation must be specified that follows the same temporal evolution as (8), and whose effect is to guarantee that (9) holds. Assuming that the initial conditions obey (9) (as they should), a way to do this is to guarantee that the “net probability flux” (what would be thought of as the mass flux if this were a problem in physics) across the point $x=0$ is always zero.

We fix a point x in one dimension and consider the probability flux across that point for both directions. By integrating all possible increasing transitions that cross this

barrier, and subsequently all possible decreasing transitions that cross the same barrier, and then adding them together, we arrive at the following requirement that the net probability flux be zero. This constraint is referred to in the physics or stochastic processes literature as a *reflecting barrier*.

$$f(0;t)M(0;t) - \frac{1}{2} \frac{\partial}{\partial x} f(x;t)V(x;t) \Big|_{x=0} = 0, \quad t > 0 \quad (10)$$

C. Initial Conditions

The functions $M(x;t)$ and $V(x;t)$ represent the first and second moments, respectively, of the rate at which the length of the queue is changing at time t , given that its current state is x . In a queuing system where the arrival process is independent of the service process, then with the possible exception of $x=0$ and an upper reflecting barrier, there is no reason to suspect that these functions should change across x . In such situations, it is only necessary to specify how these functions change over time. For most aviation applications, for example, one would expect $M(x;t)$ to be positive at the beginning of the day, negative at the end of the day, and perhaps with some additional cycles in between. One would expect $V(x;t)$ to be small (approaching zero) at the beginning and end of the day and something larger in between, and of course never negative. If this construction were extended to a queuing network, these functions could be derived entirely from the outputs $\{f_i(x;t)\}$ of upstream queues, with some time lags and with some rules for mixing them together.

Although we explicitly prevent negative queues, it also makes sense to preclude initial conditions that would seem in conflict with this goal. Thus, we require that

$$M(0;t) \geq 0 \quad \forall t \quad (11)$$

At any node to which this method is applied, one can imagine that $M(x;t)$ will be computed as the differential of the difference between the arrival rate, which we might get from the outputs of upstream processes, and the departure rate, which is related to the capacity of the airport or other resource. This being the case, (11) simply prevents an airport from serving traffic that does not exist.

At some airports, however, the rate of queue growth might depend on its current state. For example, if the total capacity of the airport is divided between arrivals and departures, and the airport has some control over that split, then in cases when there is an excess of arrivals, the airport might choose to emphasize arrivals over departures to ameliorate this queue. This is tantamount to a temporary increase in the arrival capacity of the airport. If this were repeatable and quantifiable behavior, that could be captured in differences in $M(x;t)$ across different values of x .

We must specify an initial queue length distribution. For real airport problems, the queue is empty at the beginning of the day, so one might require:

This research is supported by the National Aeronautics and Space Administration

$$f(x;0) = \delta(x), \quad (12)$$

where $\delta(\square)$ is the Dirac delta function. Alternatively, one might consider analyzing a problem starting at some other point in the middle of the day, in which case the restriction (12) is not required. At all times, however, f must be a proper density function:

$$f(x;t) \geq 0 \quad \forall x, t \quad (13)$$

$$\int f(x;t) dx = 1 \quad \forall t \quad (14)$$

III. NUMERIC SCHEME

In order to solve a system including partial differential equations and their associated boundary and initial conditions, a numerical scheme is necessary to convert that continuum problem into some discrete form appropriate for solution by computer [6]. In this paper we present a discretization method based on the well-known finite element method (FEM) that is appropriate for our problem. The construction of numeric schemes for PDEs is very much an art, and certainly a host of other schemes could be attempted, including finite difference methods.

The FEM scheme developed for this problem consists of transforming the governing differential equation with its boundary and initial conditions into linear algebraic equations that can be solved at every time step. This transformation is possible by constructing a discrete approximation to the queue length density function $f(x;t)$ using the N Lagrange basis functions ϕ_1, \dots, ϕ_N . Each basis function has a triangular shape; the collection of them is illustrated in Fig. 3 for $N=4$. Mathematically, the basis functions can be represented as follows:

$$\begin{aligned} \phi_1(x) &= \begin{pmatrix} l_2 - x \\ l_2 - l_1 \end{pmatrix}_+, \\ \phi_j(x) &= \begin{pmatrix} \min\left\{\frac{x-l_{j-1}}{l_j-l_{j-1}}, \frac{l_{j+1}-x}{l_{j+1}-l_j}\right\} \\ + \end{pmatrix}, \quad j = 2, \dots, N-1, \\ \phi_N(x) &= \begin{pmatrix} x - l_{N-1} \\ l_N - l_{N-1} \end{pmatrix}_+. \end{aligned}$$

The approximation for f can then be expressed using these basis functions as:

$$f^{L,N}(x;t) = \sum_{j=1}^N a_j(L\Delta t) \phi_j(x), \quad (15)$$

where L is the time step, N is the number of Lagrange basis functions, and $\{a_j\}$ are the parameters of the approximation.

Using the finite element method, the “solution” of the problem essentially amounts to determining the values $\{a_j\}$.

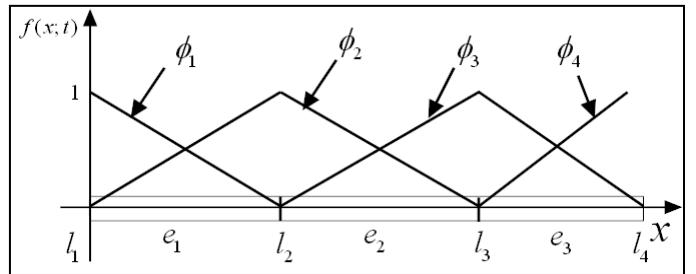


Fig. 3. Lagrange basis functions for the finite element method

The left hand side of the PDE (8) can now be approximated by:

$$\frac{\partial f(x;t)}{\partial t} \square \frac{f^{L+1} - f^L}{\Delta t},$$

and the dynamics can be re-written as:

$$\frac{f^{L+1} - f^L}{\Delta t} = \frac{1}{2} \frac{d^2}{dx^2} (V^{L+1} f^{L+1}) - \frac{d}{dx} (M^{L+1} f^{L+1}) \quad (16)$$

We enforce (16) by defining the residue r , which is essentially the difference between the LHS and RHS of (16),

$$r = \frac{1}{2} \frac{d^2}{dx^2} (V^{L+1} f^{L+1}) - \frac{d}{dx} (M^{L+1} f^{L+1}) - \frac{f^{L+1} - f^L}{\Delta t}$$

We force that residue to zero by using a test function $w(x)$. We equate all of the projections of the residue on w to be zero; i.e., $\int_{\Omega} r w dx = 0$, where Ω is the domain of interest in x and $\partial\Omega$ its boundary. Integrating by parts yields:

$$\begin{aligned} \frac{1}{2} \int_{\Omega} \frac{d}{dx} (V^{L+1} f^{L+1}) \frac{dw}{dx} dx - \int_{\Omega} M^{L+1} f^{L+1} \frac{dw}{dx} dx + \\ \int_{\Omega} \frac{f^{L+1}}{\Delta t} w dx = \int_{\Omega} \frac{f^L}{\Delta t} w dx + \\ \left[\frac{1}{2} \frac{d}{dx} (V^{L+1} f^{L+1}) - M^{L+1} f^{L+1} \right] w \Big|_{\partial\Omega} \end{aligned} \quad (17)$$

where the last term on the RHS depends on the boundary conditions.

We assume that the interval is closed, and that at the right boundary $x = l$, we would like the net probability flux to be 0. For some large l , the probability density function will approach 0 for all $x > l$. This will make the net probability flux approach zero at $x = l$, although it cannot be absolutely guaranteed. This is discussed more in the conclusions. Together with equation (10), we conclude:

$$\left[\frac{1}{2} \frac{d}{dx} (V^{L+1} f^{L+1}) - M^{L+1} f^{L+1} \right] w \Big|_{\infty} = 0.$$

We parameterize the test function w with the Lagrange basis functions $\{\phi_i\}$ and parameters $\{b_i\}$:

$$w(x) = \sum_{i=1}^N b_i \phi_i(x) \quad (18)$$

We use the Lagrange approximations of w and f to obtain:

$$\sum_{i=1}^N b_i \left[\sum_{j=1}^N a_j^{L+1} K_{ij} - R_i \right] = 0 \quad \forall \{b_i\} \quad (19)$$

where

$$K_{ij} = \frac{1}{2} \int_{\Omega} V^{L+1} \phi_j' \phi_i' dx - \int_{\Omega} M^{L+1} \phi_j \phi_i' dx + \frac{1}{\Delta t} \int_{\Omega} \phi_j \phi_i dx$$

$$R_i = \frac{1}{\Delta t} \int_{\Omega} \phi_i \left(\sum_{j=1}^N a_j^L \phi_j \right) dx$$

In the last two equations, we denote $a_j^L = a_j(\Delta t)$ and suppress the dependence of the basis functions $\{\phi_i\}$ on x for the sake of clarity. As mentioned before, we have also assumed that the function $V(x; t)$ is constant in x .

Since the set $\{b_i\}$ is arbitrary, (19) is equivalent to solving the linear algebraic equations:

$$\sum_{j=1}^N a_j^{L+1} K_{ij} = R_i \quad \text{for } i = 1, 2, \dots, N \quad (20)$$

The solution of (20) is the set of parameters $\{a_j\}$ which define $f(x; t)$ according to (15). One of the advantages of the finite element method is the ability to solve these algebraic equations element by element. The N Lagrange basis function approximation defines $N-1$ elements, which makes it possible to solve $N-1$ independent algebraic equations.

IV. MODEL VALIDATION AND RESULTS

In this section, we show some results of applying the modeling with different input data sets. First, it is important to acknowledge that there are ranges of the input data that can lead to numerically unstable results. It turns out that scaling the inputs can avoid this artifact, but more investigation is required to determine exactly what circumstances are most susceptible and how to avoid them deliberately. Fig. 4 shows, for example, the kind of results that can be expected with data that lead to numerical instability. The oscillating triangular-shaped curves are generally indicative of this result.

Fig. 5 shows the results for an input data set that consisted of a mean vector \mathbf{M} representing the function:

$$M(x; t) = \begin{cases} 0.2, & \text{if } t \leq 14 \\ -0.2, & \text{otherwise} \end{cases} \quad \forall x > 0$$

The variance was modeled as a constant value equal to $V(x; t) = 0.1$ for all x and t . By contrast, the data for Fig. 4 were identical except that the mean values were scaled by a factor of 9. It will be shown later that it is not just the scaling of the means that matters, but their scale relative to the scale of the variance. The values chosen for the inputs are somewhat arbitrary; the main goal was to produce well-behaved outputs that would show a pattern that was consistent with the inputs.

In the upper left graph of Fig. 5, it is obvious that the mean and variance of the queue length distribution are growing up to time slice 14. This is confirmed in the lower two graphs. The upper right graph is a zoomed in region of the solution space showing only the first 20 time slices, where the change in behavior from a growing queue length to a declining one is more apparent. In the lower right graph, the variance starts to decline some time after time slice 14, as there is a natural delay in which a reduction in the mean input to the system can be manifested as a change in the variance of the output.

Because scaling of the input values is one mechanism by which one can avoid numerical instability, it is interesting to observe the impact scaling can have on the outputs. Fig. 6 shows results similar to Fig. 5, except that the mean inputs have been scaled by a factor of 10 and the variance inputs by a factor of 100. Notice that the mean and variance profiles that result are remarkably similar. When testing with a variety of scaling factors, it turns out, not surprisingly, that the equilibrium mean queue length is linear in the scale factor applied to the mean, and the equilibrium queue length variance is linear in the scale factor applied to the variance, when that is the square of what was applied to the mean. Thus, one important conclusion is that the model is invariant, other than a variable transformation, to scaling, which means that if different scales are more adept at avoiding numerical instability, this is a good recipe to use for that purpose.

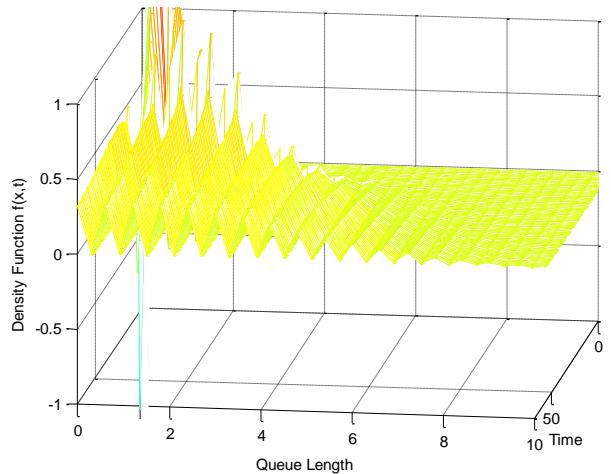


Fig. 4. Example of numerical instability

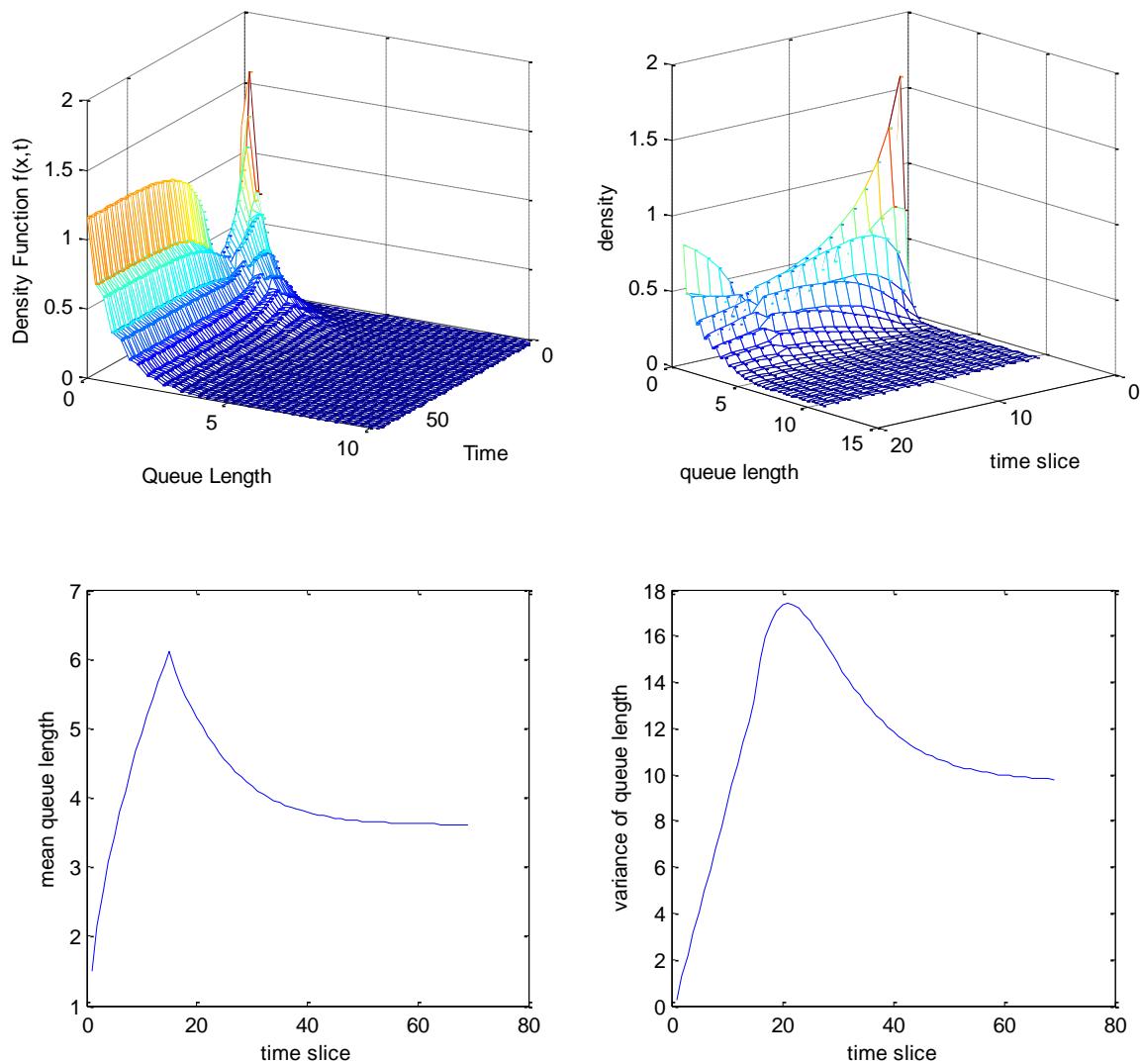


Fig. 5. Diffusion model results with constant variance and two-stage mean

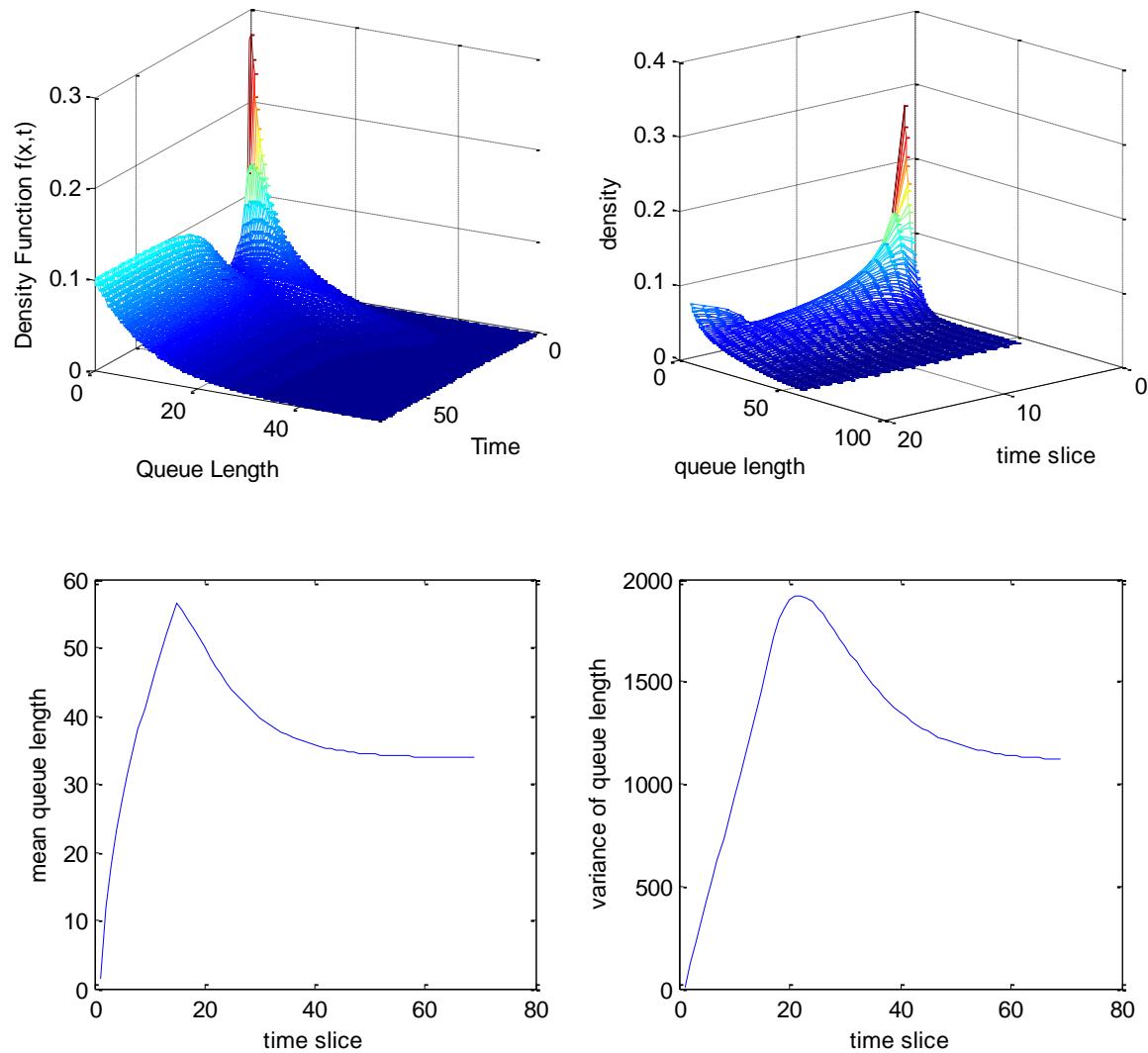


Fig. 6. Diffusion results scaled by 10 in the mean and 100 in the variance

V. CONCLUSIONS

The paper has presented the mathematical construction of a continuum approximation to a queuing system that might represent a single congested resource in the National Airspace System, such as an airport, a runway, or some en route resource. The result is derived from the diffusion approximation. A numeric solution scheme based on the finite element method is also shown.

The use of this type of approximation requires one to be comfortable with some of the assumptions made in the paper, such as the willingness to consider non-integer queue lengths. That notwithstanding, the method has seen considerable application in other areas of queuing theory that also deal with countable objects, so this assumption is not unique to the aviation context.

This result is a stepping stone in what will hopefully be a larger system of inquiry into the use of such continuum approximations to study systems of aviation queues. In particular, the ability to model the propagation of both the mean and the variance of delay statistics through a connected network would mark a major leap forward in the performance analysis of the aviation system.

Some other detailed analyses are probably in order first, however. First, while this paper shows evidence that numerical instability can occur, and showed examples of how input scaling was used to avoid this deleterious effect, this is only a brief empirical presentation, and a more thorough investigation should be conducted of the conditions that give rise to this instability and some ideas about how to prevent it more systematically. For example, it is probably the case that

for cases where the queue is assumed to be initially empty, because the bulk of the density function is concentrated near the origin, a finer mesh spacing would be appropriate in that region in the early time slices to give the problem the ability to better distinguish that fine resolution. Some formal mathematical results are available for various types of PDEs evaluated numerically using FEM; these results should be studied and adapted to the problem at hand, if possible.

The second necessary step would be to conduct a thorough validation effort. One would expect, for example, to be able to replicate the known steady-state results from that small set of queuing systems for which equilibrium results are known in closed form. Furthermore, a Monte Carlo exercise could be conducted for a number of other cases that cannot be found analytically.

REFERENCES

- [1] Newell, G.F., Applications of Queueing Theory, Chapman-Hall, 1971.
- [2] Kobayashi, H., “Application of the diffusion approximation to queuing networks II: Nonequilibrium distributions and applications to computer modeling,” J. ACM, vol. 21, no. 3, pp. 459-469, July 1974.
- [3] Gaver, D.P., Jr., “Diffusion approximations and models for certain congestion problems,” J. Appl. Prob., vol. 5, no. 3, pp. 607-623, December 1968.
- [4] Kimura, T., “Diffusion approximation for an M/G/m queue,” Oper. Res., vol. 31, no. 2, pp. 304-321, 1983.
- [5] Kimura, M., “Diffusion models in population genetics,” J. Appl. Prob., vol. 1, no. 2, pp. 177-232, December 1964.
- [6] Pepper, D.W. and J.C. Heinrich, The finite element method: Basic concepts and applications, Taylor & Francis, 1992.

Track 2

Airline Operations and Marketing

The Responses of Traditional Airlines to Low Cost Airlines

Panarat Srisaeng

School of Aerospace, Mechanical and Manufacturing Engineering
Royal Melbourne Institute of Technology
Melbourne, Australia

Panarat.srisaeng@rmit.edu.au

Abstract — The significant growth in low-cost carriers has generated a lot of competition in the airline industry. Traditional airlines have come under pressure with a collapse in profitability while many low-cost carriers have enjoyed profits. The low cost airline business model can cut costs to 40-50% of traditional airline costs while the traditional airline business model has been challenged to cover costs. Major airlines have attempted to reshape their business model to boost revenue, cut costs and react to competitive threats from low-cost carrier. These strategies include establishing their own low cost airlines, increasing labour efficiency, intimate low cost airline operation, and introducing charges for catering and luggage. This paper aims to investigate the responses of traditional airlines to low-cost carriers. Secondary sources of information for this paper will include a review of current academic literature and published industry sources.

Keywords- *low-cost carriers; airline; competition*

I. INTRODUCTION

Over the past two decades, the traditional network model for scheduled airline service has delivered market growth and in good years, modest profits (Morrell, 2005). However, with the emergence of low cost carriers the aviation industry has changed. The Low Cost Carrier Model is a revolutionary change to the airline business which threatens traditional airlines. The challenge being that low cost carriers attract passengers with a simple business model, low fares; in turn these lower fares then increase air travel demand. Furthermore, while the low cost carrier's target market is the leisure traveler, due to simple economics some business travelers also turn to the low cost carrier. In this way, low cost carriers not only generate new airline passengers, those who would not otherwise travel by air, but also steal passengers from traditional airlines.

As a result low-cost airlines have enjoyed profits and growth while traditional airlines' profitability has collapsed (Dennis, 2007). This situation challenges traditional airlines to revisit their business model, forcing them to reinvent themselves. Traditional carriers' reaction to the new situation is crucial as it increases the airline industry's competitive environment. This paper aims to investigate the responses of traditional airlines to low cost carriers. Strategies include

significant cost cutting, establishing their own low cost airlines, outsourcing, intimating low cost airlines operation, and instituting new charges for catering and luggage.

II. REMOVE SIGNIFICANT COST

A. *Airline Operating Costs*

Airline operating costs can be divided into two main categories, direct and indirect. Direct operating costs consist of flight operations, maintenance cost depreciation as shown in this following table.

TABLE 1: AIRLINE OPERATING COST CATEGORIES

<i>Direct Operating Costs</i>
1. Flight operations: <ul style="list-style-type: none"> • Flight crew salaries and expenses • Fuel and oil • Airport and en-route charges • Aircraft insurance • Rental/lease of flight equipment/crews
2. Maintenance and overhaul: <ul style="list-style-type: none"> • Engineering staff costs • Spare parts consumed • Maintenance administration
3. Depreciation and amortisation: <ul style="list-style-type: none"> • Flight equipment • Ground equipment and property • Extra depreciation • Amortisation of development costs and crew training
Indirect operating costs
4. Station and ground expenses: <ul style="list-style-type: none"> • Ground staff • Buildings, equipment, transport • Handling fees paid to others
5. Passenger services: <ul style="list-style-type: none"> • Cabin crew salaries and expenses • Other passenger service costs • Passenger insurance
6. Ticketing, sales and promotion
7. General and administration
8. Other operating costs

Source: Doganis,2002

- Direct Operating Costs

There are five items in flight operation costs: flight crew, fuel, airport charges, aircraft insurance, and leasing costs. Flight crew costs include direct salaries, travelling and stopover expenses, as well as allowances, pensions, insurance and any other welfare payments. Cockpit crews' salaries depend on aircraft type, as in general the larger the aircraft the higher the salaries are. Flight crew costs can be calculated on a route-by-route basis or expressed as an hourly cost per aircraft type. Therefore, the total flight crew costs for a particular route or service can be calculated by multiplying hourly flight crew costs of the aircraft being operated on a route, by the block time for the route. For fuel costs, this cost element is aircraft specific. Fuel consumption varies by aircraft type, number and size, or thrust of engines and type and age of those engines. During operations, actual fuel consumption varies by sector length, aircraft weight, wind conditions, cruise altitude, and so on. Fuel consumption is usually calculated based on the number of engines on the aircraft flying the route multiplied by the hourly consumption for that engine and by block time. In addition fuel costs also include fuel charges levied by some airport authorities and government levied fuel taxes. Next cost item is airport charges which have two elements a landing fee and a passenger charge. The landing fee relates to maximum take-off aircraft weight and a passenger charge levied on the number of passengers boarding at an airport. Some airports collect the fee directly from each passenger on departure, which does not appear as an airline cost. Airlines also have to pay en-route navigation charges to cover the cost of en-route navigation aids their aircraft use while flying. This charge relates to aircraft weight and distance flown over a country's air space. For aircraft insurance, the annual insurance premium is calculated as a percentage of the full purchase price which may be between 1.5 per cent and 3 per cent. The annual premium can be converted into an hourly insurance cost by dividing it by the total number of expected block hours during the year. The last item in this category is the rental/ lease of flight equipment/crews. Leasing aircraft is increasingly widespread among airlines. There are two types of lease: operating and financial leases. Operating leases are generally five years or less then after leasing aircraft ownership remains with the lessor, while financial leases are 10 years or more, which after ownership transfers to the airline. The second category for direct operating costs is maintenance and overhaul. Maintenance costs include both routine maintenance and maintenance checks carried out between flight or overnight, but also more extensive periodic overhauls and major checks. Major costs here are the cost of maintenance staff and spare part consumption. Depreciation and amortization are the last cost element in this cost category which includes amortization and depreciation of capital leases, office/flight/ground station equipment and other fixed assets (Doganis, 2002).

- Indirect operating costs

Indirect operating costs include station and ground expenses, passenger services costs, as well as ticketing, sales and promotions, and general and administrative costs. The first

groups of indirect operating cost are station and ground costs involving airline services provided at an airport. These include salaries and expenses of all airline staff at the airport base, lounges, ground handling equipment, ground transport and office equipment. The second item for indirect operating cost is passenger services, passenger services costs can also be divided into three groups. The first group is pay, allowances and expenses related to aircraft cabin staff such as hotel and other costs associated with overnight stops. The second group is costs directly related to passengers for instance in-flight catering, transit passengers' accommodation, meals, other facilities provided on the ground for passenger comfort and expenses incurred due to delayed or cancelled flights. Finally is the annual premium insurance charge for passenger liability insurance and passenger accident insurance which depends on each airline's safety record.

Ticketing, sales and promotion costs are the third item which included commission of fees paid to travel agencies for ticket sales, credit card companies, global distribution systems, as well as the cost of retail ticket offices and all promotional expenditure. Finally, General and administrative costs normally include only those cost elements which cannot be allocated to any categories (Doganis, 2002).

B. Cutting staff costs

This section will analyze cost elements traditional airlines have eliminated. Table 2 shows the distribution of total operating costs between various cost elements. As discussed before, staff costs are associated with many other cost elements so it is summarized as one cost element representing total staff costs. It is clear then that staff costs are the most significant cost element in airline operating cost followed by fuel cost and rental (Dempsey and Gesell, 2006).

TABLE 2: BREAKDOWN OF AIRLINE OPERATING COSTS 1969- 2004

Cost	1969	1973	1980	1990	1995	2004
Staff salaries & benefits	40.9	45.6	37.3	33.8	36.3	31.5
Equipment rentals	n/a	n/a	1.8	7.1	15.5	8.7
Fuel & Oil	12.4	12.1	31.0	17.7	11.6	18.0
Travel agent commissions	2.5	3.2	3.4	10.0	9.3	1.6
Food	3.6	3.9	n/a	n/a	3.4	2.0
Landing fees	1.9	2.6	1.7	1.8	2.2	2.2
Advertising and other promotions	2.9	2.4	1.7	2.1	1.7	0.8
Interest on debt	3.8	3.3	n/a	n/a	n/a	n/a
All others	32.0	26.9	21.2	27.2	20.0	35.2

Source: adapted from Dempsey (2006)

It also shows staff costs are the largest proportion of operating cost for both low cost and traditional carriers. As shown in table 3, Morrell (2005) compares airline operating costs between Southwest Airlines and US Airways. This shows Southwest's staff cost as 39% and US airways' as 40.8% which differ by 1.8 %, followed by fuel and maintenance costs.

TABLE 3: OPERATING COST FOR SOUTHWEST AND US AIRWAYS

Cost Category	Southwest (%)	US Airways (%)	% point different
Staff costs	39.0	40.8	-1.8
Fuel	14.9	9.8	+5.1
Maintenance	7.6	5.1	+2.5
Sales commissions	1.1	1.6	-0.5
Landing/rents	6.8	5.4	+1.4
Aircraft rent/ depreciation	10.6	10.3	+0.3
Other	19.9	27.0	-7.1
Total	100.0	100.0	

Source: Compiled from airlines annual reports and Morell, 2005

Most costs categories depend on external environment for example, fuel and oil, interest rates, landing fees or aircraft costs which are difficult to control. It seems staff costs are potentially the most controllable operating costs. Traditional airlines focus closely on staff cost elements and service costs to cut airline operating costs (Doganis, 2001). So, controlling staff costs is seen as key for airlines success.

Staff costs typically range between 30-40% of operating costs which is the biggest single airline expense. Staff cost elements include salary, benefits, payroll taxes for management and any associated social charges. According to several airlines, management and administrative staff often account for about 10% of labour costs, pilots 31-35%, flight attendants 13% and mechanics about 13-16% (McCartney, 2002). Traditional airlines have tried to reduce staff costs through salary and benefit reductions, and productivity improvements. Several legacy airlines have attempted to freeze or reduce salaries or benefits, employ new staff on less general terms and conditions or even lay off staff. According to the Association of Flight Attendants there are about 100,000 flight attendants in the United States down from about 125,000 in 2000 and their income has decreased by 20 percent (Higgins, 2008).

In 1994, Delta Airlines planned to reduce its operating cost of flying a seat one mile (available seat mile or ASM) from 9.26 cents to 7.5 cents by 1997, by slashing its costs 19% over three years. Delta achieved this target by using aggressive restructuring plan to eliminate 20% of its work force, outsourcing, and reducing staff benefits (O'Brian, 1994). Delta then eliminated 4,500 full time customer service employees, and in 2005 announced it would outsource nearly half its major airline maintenance and overhaul work. This contract aimed to save \$240 million a year and cut its heavy maintenance costs by 34% over five years but 20% of staff would lose their job (Field, 2005). These changes ruined the Delta corporate culture of labour management harmony, service levels dropped sharply, hundreds lost luggage and angry passengers abounded (Dempsey and Gesell, 2006).

III. OUTSOURCING

Some airlines outsource labour intensive activities such as ground handling, ticketing, catering, cleaning and maintenance services to control salaries and benefits. They have long contracted other airlines to provide activities where larger airlines have operations and which leads to economies of scale. British Airways outsource ground transport at Heathrow and Gatwick airports and sold catering department to Swissair's Gate Gourmet. In turn Aer Lingus sold its entire maintenance division to FLS engineering (Doganis, 2001). Low-wage airlines like Continental and American West have outsourced functions as such maintenance services. ValuJet outsourced heavy maintenance and reservation. United contracted out sky cap and cleaning services then sold off its catering unit to Dobbs, for \$120 million. As a result it saved \$71 million of catering renovation resulting in savings of \$320 million over 7 years. Airlines may even outsource staff from lower salary countries. For instance, Japan airlines outsourced staff from Thailand and Singapore, which are lower salary countries or from the UK and Germany which are relatively lower wage countries. A Thai flight attendant is paid about 10% of a Japanese flight attendant salary but is well paid in comparison to other jobs in Thailand. Japan Airlines offered local staff as much as \$600,000 to quit their job and stopped employing Japanese flight attendant in 1992 then replaced them with overseas staff on less favorable terms and conditions. In 1989 4% of staff were non Japanese by 1998, 28% of all staff were not Japanese (Reitman and Sapsford, 1994). Singapore Airlines and Austrian Airlines also outsourced overseas staff members while the former employed staff from Malaysia and Indonesia the latter employed accounting staff from India (Dempsey and Gesell, 2006).

IV. FUEL PRICE HEDGING PROGRAM

Fuel costs represent 10-20% of operating costs which is the second biggest cost element. It is difficult to manage this as fuel costs are an external factor. Over the past several years fuel costs have risen substantially, putting a pressure on airlines to control operating costs. In 2000 West Texas Intermediate Crude stood at \$30.30 per barrel and it increased to 63.27 per barrel by 2006.

However, airlines can protect themselves against the risk of rising fuel costs by fuel price hedging programs. A fuel price hedge program is a contractual tool where an airline commits to buy fuel at an agreed upon fixed price at some point in the future, regardless of the market price at that time. If the market price is above the agreed upon fixed price, the buyer gains. If the market price is below the agreed upon fixed price, the buyer loses (Barton, 2008). Table 5 shows fuel expenses and hedging strategies for US domestic airlines.

TABLE 4 PRICE OF WEST TEXAS INTERMEDIATE CRUDE OIL

Year	Price per Barrel (\$USD)
2000	30.30
2001	25.92
2002	26.10
2003	31.14
2004	41.44
2005	56.48
2006	63.27

Source: Air Transport Association of America

From table 5, in 2003 fuel cost averaged over 16% of total operating costs for US domestic airlines. Southwest airlines and JetBlue were industry leaders in fuel hedging with 82% and 40% of expected 2004 fuel consumption hedged as of December 2003, both airlines stated fuel hedging is key to their low-cost strategy and believe this strategy forms a competitive advantage. Across 2001 – 2003, Southwest cut its annual fuel costs by \$171 million, \$45 million, and \$80 million, respectively, through its fuel hedging program. Like Southwest, JetBlue managed their fuel costs by using fuel hedging program. In 2002-2003 JetBlue reduced its annual fuel costs by \$4 million and \$1 million respectively (Cobbs and Wolf, 2004). While some major airlines did not use fuel hedging program such as American, United or Northwest. These airlines risk taking rising fuel price into their business model. They pass fuel costs on to passengers by adding fuel surcharges to airfares. However, when fuel prices rise dramatically airlines cannot pass all of the cost on to passengers (Zea, 2002).

Fuel hedging programs have several advantages. Firstly, hedging airlines can better predict future expense and earning, which help increase financial market confidence. Secondly, hedging lets airlines take advantage of investment opportunities when fuel prices are high. Carter and Simkins (2002) show measurable fuel hedging by airlines can increase the value of the firm an estimated 12-16%. There is therefore a positive correlation coefficient between airline valuation and the airline's fuel hedging levels.

TABLE 5: THE FUEL EXPENSE AND HEDGING STRATEGIES FOR THE US DOMESTIC AIRLINES

<i>Company</i>	<i>ASM FY 2003 (in Millions)</i>	<i>Fuel as a % of Operation costs FY 2003</i>	<i>Avg % of Fuel Hedge FY 2004</i>	<i>Avg % of Fuel Hedge FY 2005</i>
Airtran Holdings	10,046	21.5	35	12
America West	23,373	16.4	11	0
American	165,209	15.2	12	4
ATA	21,126	19.2	0	0
Continental	78,385	14.5	0	0
Delta	134,000	13.8	32	0
Frontier	2,841	17.9	7	0
JetBlue	13,639	17.8	40	0
Midwest Air	2,968	19.6	0	0
Northwest	88,593	15.9	0	0
Southwest	71,790	15.2	82	60
United	136,630	13.7	0	0
US Airways	58,106	11.7	30	5

Source: Company SEC filing and Carter (2002)

V. CHARGE FOR CATERING AND LUGGAGE

Some network airlines cut their costs allowing them to lower their airfare by no longer offering a free meal. Cutting all catering reduces both direct and indirect costs. When airlines have no catering service they can reduce turnaround time as the aircraft does not need to be cleaned and catered. They can also gain more seats when galley space is replaced by seats. Further, cabin staff can be reduced to the safety minimum. The US Airways Group was the first major American airline to charge for coffee and sodas, although the Association of Flight Attendants objects to collecting the \$1 and \$2 fees for non-alcoholic drinks. United Airlines no longer offers a free meal on short-haul economy flights. However, fresh food menu or snack box options are available for purchase \$7USD and \$5USD each. Alcoholic beverages are also available for purchase.

Low-cost carrier success has forced a revaluation of short-haul product by traditional airlines. Low cost airlines either offer no catering or a basic paid-for-service and their airfares are only half or less of the network airlines. Then, suddenly free catering became the most visible symbol of difference between operators. Traditional airlines have gradually reduced economy class free meals. This not only saves money but increases business class product differentiation. The argument being that no-one buys an air ticket because of the food. So if ticket prices can be cut through cutting out food, commercial success will follow. However, the danger for traditional airlines

is that they can never match cost levels and average fares of low cost airlines. If inclusive economy class catering is eliminated, passengers may then see no reason for using these airlines. Dennis (2008) argues legacy carriers have rushed to strip out catering provision on short-haul flights. They believe the Frequent Flyer Program is the only frill valued by passengers. Most of these airlines have performed very poorly, not helped by negative passenger perception coming from no in-flight service, disillusioned staff and fares often higher than Southwest or Jet Blue

In the United States, in May 2008, five of the six major airlines started charging passengers up to \$25 for a second bag. This new fee was levied by Continental, Delta, Northwest, United and US Airways. American Airlines announced that it would charge a \$15 fee for the first checked bag, on top of the \$25 second bag. United Airlines and US Airways then applied this fee as well. While Delta did not join the other major airlines charging for the first bag, it doubled the fee for a second bag from \$25 to \$50. United estimated that new baggage fees would generate about \$275 million revenue a year. Although full service airlines in the US attempted to charge passengers more and more, Southwest, the largest low fare US airline, did not join other airlines in charging fees for previously free features, such as checking bags, and turned its decision to forgo them into a marketing campaign. "Bags fly free," the airline declares on its Web site that passengers can check two bags free, and must pay for the third. Gary D. Kelly, Southwest's chief executive said the airline remains reluctant to add baggage charges, even though it is studying whether to impose other fees, which the industry calls "ancillary revenue." This situation even gives it a big advantage over those airlines to gain more passengers (Maynard, 2008).

VI. ESTABLISH A SUBSIDIARY LOW COST CARRIER

In one competitive response to the growth of new entrants, traditional airlines establish their own carriers using the low-cost no-frills business model. Of many attempts to set up a no-frills low-cost carrier as a subsidiary of a traditional airline however, most have failed. Table 6 shows an overview of both inactive and active low-cost subsidiary airlines where their mainline also operate both a full-service network carrier and low cost carrier business model.

Continental Airlines, the US' fifth largest carrier, established its subsidiary low-cost carrier Continental Lite to compete in the low cost carrier market. Before it closed down Continental Lite offered low-cost flights, primarily east of the Mississippi River. Continental then changed its pricing structure, moving away from many of the bargain-basement fares which had cut into the airline's profitability. In 1994, Continental Airlines suffered a monthly loss of about \$55 million, of which up to 70% could be attributed to Continental Lite. Kevin C. Murphy, an airline analyst at Morgan Stanley, stated that Continental Airlines is a business where trying to do one thing well is difficult enough (Bethune and Huler, 1998).

United Airlines announced low cost service called Shuttle by United in 1994 to compete primarily with Southwest Airlines. Shuttle operated along the West Coast of the US and offered fares as low as \$62 for every seat on every flight, some

of which are comparable to Southwest's fares. But the Shuttle's customer received some frills, like a seat assignment at the airport and the right to earn mileage in United's frequent flier program. Shuttle eventually shut down in 2001 because it could never get costs low enough. After Shuttle, United Airlines attempted to launch another low-cost airlines, Ted. Ted was established in November 2003 based in Denver, serving 23 destinations in the United States and Mexico with 57 Airbus A320 aircrafts 156 seats with all economy class. Ted served the market for 5 years, and was shut down in June 2008 due to the jet fuel crisis.

TABLE 6: CLOSED DOWN AND ACTIVE LOW COST AIRLINES IN THE SAME AIRLINES GROUPING

Closed down		Active	
Airline Grouping	Low-cost unit	Airline Grouping	Low-cost unit
Continental Airlines	Continental Lite	British Midland	Bmibaby
United Airlines	Shuttle by United	KLM	Transavia (Basiq Air)
United Airlines	Ted	Lufthansa/Euro wings	German wings
Delta Airlines	Delta Express	Qantas Airways	JetStar
Delta Airlines	Song	Qantas Airways	JetStar Asia
US Airways	MetroJet	Japan Airlines	JAL express
KLM	Buzz	Thai Airways	Nok Air
British Airways	Go	Singapore Airlines	Tiger Airways
Lufthansa	Lufthansa express		
Frontier Airlines	People Express		
Austrian Airlines	"Austrian Bratislava"		
Air Canada	Zip, "Tango" 2004		
SAS	Snowflake		
Finnair/Nordic Air link	Fly Nordic		
LOT	Central wings		
Qantas Airways	Australian Airlines		
Air New Zealand	Freedom Air		

Source: Compiled from Graf (2005) and Airlines' website

Delta Airlines launched Delta Express as no-frills airlines in 1996 based in Orlando international airport serving 31 domestic markets, then replaced it with Song in 2003. Song was a subsidiary of Delta Airlines targeting leisure passengers, flying to 21 destinations in the United States and Caribbean. Delta discontinued Song in 2006.

British Airways set up GO in November 1997 based in London Stansted to compete in the European low-cost market, dominated by Ryanair easyJet and Debonair. In 2002 it was eventually taken over by easyJet with the reason given being that BA wanted to focus on the business it understood best, that of a full-service carrier.

The idea of running two different and actually conflicting airline business models simultaneously often leads to poor quality, dissatisfied customers, and discouraged employees (Porter, 1996). Many subsidiary low cost airlines' cost structure is the same as the mainline operation, which means lower fares might not even cover costs. One critical factor seems to be the degree of independence the low cost operation is given by the mainline operation. If a subsidiary low cost carrier is given an independent management it can build its own cost structure and business plan. The low cost airline can generate substantial cost savings by separate labour contracts, choose its own distribution channels even develop an independent network and timetable (Dietlin, 2004). However, when highly independent management is permitted cannibalization is inevitable. The low cost operation then competes directly with the mainline operation, since the low cost carrier operates point to point routes which are likely the same as the network carrier. Since a network airline has a very dominant position in the market the low cost airline will start cannibalizing mainline operation in these markets. This was one of the main reasons why British Airways sold its low cost subsidiary airline GO (Dietlin, 2004).

VII. INTIMATE LOW COST CARRIER OPERATION

Some traditional airlines in the US or Europe intimate economy class operations of low cost airlines. For instance, Swiss International Airlines changed its economy class on all its European flights in 2003 to make it more similar to the low cost carrier product. The Swiss economy class fare was reduced to match that of the low cost carrier. Booking online was introduced and only provided a purchased meal. This strategy let Swiss International Airlines' load factor increase 16 percent while revenue per ASK rose 3.2 percent, however yield decreased 11 percent (Dietlin, 2004).

One aspect of concern regarding the conversion of the short haul economy class product to a low cost offering is the seat availability of connecting passengers. The low cost operation which leads to a lower fare stimulates demand for travel. However where an airline is part of a global network with connecting flights between short haul and long haul flights, it might be that new passengers displace connecting traffic to the airline's long haul flights. If these connecting passengers do not obtain seats on the short haul connecting flight they will also miss the airline's long haul flight. This is very important because long haul flights may not maximize revenue and this will jeopardize mainline operation yield (Dietlin, 2004).

VIII. CONCLUSION

The competitive strategies adopted by the traditional airlines in reaction to the competitive threat from low cost airlines included setting up a low cost carrier subsidiary, introducing charged for catering and luggage, reducing staff cost and intimating low cost carrier operation. Reducing staff cost and improve staff productivity are successful strategies for the traditional airline because staff cost is the great proportion in airline operating cost. This strategy causes a little negative impact to passengers. While setting up a low cost subsidiary is

a very inefficiency option because running two brands under the same umbrella leads to incompatibilities of business management.

REFERENCES

- [1] BARTON, H. (2008) How to think about hedging. Purchasing. MA, Reed Business Information.
- [2] BETHUNE, G. & HULER, S. (1998) From Worst to First : Behind the Scenes of Continental's Remarkable Comeback, New York, Wiley.
- [3] CARTER, R. & SIMKINS (2002) Does Fuel Hedging Make Economics Senses? The Case of the Us Airline Industry.
- [4] COBBS, R. & WOLF, A. (2004) Jet fuel hedging strategies: Options available for airlines and a survey of industry practices. Kellogg Northwestern University.
- [5] DEMPSEY, P. S. & GESELL, L. E. (2006) Airline Management for the 21st Century, Coast Aire Publications, L.L.C.
- [6] DENNIS, N. (2008) End of Free Lunch? The response of traditional European airlines to the low-cost carrier threat. Journal of Air Transport Management, 13, 311-321.
- [7] DIETLIN, P. (2004) The Potential for Low-Cost Airlines in Asia. Department of Civil and Environmental Engineering. Massachusetts, USA, The Massachusetts Institute of Technology.
- [8] DOGANIS, R. (2001) The Airline Business in the Twenty -first Century, London and New York, Routledge.
- [9] DOGANIS, R. (2002) Flying off course: The economics of international airlines, Routledge.
- [10] FIELD, D. (2005) Delta moved to outsource. Airline Business. London.
- [11] GRAF, L. (2005) Incompatibilities of the low-cost and network carrier business models within the same airline grouping. Journal of Air Transport Management, 11, 313-327.
- [12] HIGGINS, M. (2008) Flying the Unfriendly skies. The New York Times. New York.
- [13] MAYNARD, M. (2008) Southwest Turns a Profit for 69th Straight Quarter. The New York Times. New York.
- [14] MCCARTNEY, S. (2002) Why Airlines Focus on Labour Costs For Cuts; Then There's the Food. Wall Street Journal New York.
- [15] MORELL, P. (2005) Airlines within airlines: an analysis of US network airline response to low cost carrier. Journal of Air Transport Management, 11, 303-312.
- [16] MORRELL, P. (2005) Airlines within airlines: An analysis of US network airline responses to Low Cost Carriers Journal of Air Transport Management, 11, 303-312.
- [17] O'BRIAN, B. (1994) Delta Air to Pare up to 15,000 jobs, or 20% of staff in big restructuring. Wall Street Journal. New York.
- [18] PORTER, M. E. (1996) What is strategy? Harvard Business Review, 74, 61-78.
- [19] REITMAN, V. & SAPSFORD, J. (1994) To see issues Vexing Japanese Business now consider JAL flight 76. Wall Street Journal. New York.
- [20] ZEA, M. (2002) Is Airline Risk Unmanageable? , Mercer on Travel and Transport.

Estimating Domestic U.S. Airline Cost of Delay based on European Model

Abdul Qadar Kara(Ph.D. Candidate), John Ferguson(Ph.D. Candidate), Karla Hoffman(Ph.D.), Lance Sherry(Ph.D.)

George Mason University, Center for Air Transportation Research

4400 University Drive, Fairfax, VA, USA

akara@gmu.edu;jfergus3@gmu.edu;khoffman@gmu.edu;lsherry@gmu.edu

Abstract— Researchers are applying more holistic approaches to the feedback control of the air transportation system [12-13]. Many of these approaches rely on economic feedback, including the cost of delays to the airlines. Therefore, finding the true cost of a delay is essential for air transportation management. A 2004 EuroControl study [2] describes a methodology and presents results detailing the cost to airlines of delays during various segments of a trip. The costs are divided into short delays (less than 15 minute) and long delays (greater than 65). The data used in the study consisted of data collected from European airlines, air traffic management as well as interviews and surveys conducted by the research team. However, their model is not explicitly defined and therefore no sensitivity analysis is possible in case the involved cost factors change significantly (e.g. fuel). Furthermore, the model is generated based on data from EU airlines for only 12 aircraft, so applying these delay costs to other aircraft or US airlines is not possible. This paper details a method for applying these delay costs to other aircraft and other airlines. The individual cost factor delays are applied to US data. The approach allows one to update the cost whenever any of the factors (crew, fuel, maintenance, and ground costs) change. It considers the size of the aircraft when making such calculations, both from the perspective of fuel burn and passenger costs. Data for Philadelphia airport (PHL) is displayed as a case study to show current delay costs.

Keywords-component; *airline delay costs; airline delays; economic modeling of airlines;*

I. INTRODUCTION

The airline industry moves millions of passengers and tons of cargo annually. The Schumer report estimated that in 2007, airport delays cost about 40.7 billion dollars to the economy [1]. Disruptions in one part of the airspace impact the entire network as delays propagate. It is estimated that almost 50% of the entire airspace delays are caused by delays that originate at the New York/New Jersey/Pennsylvania airports.

This implies that delays and their true costs are vital to airport and airspace management decision making. Similarly researchers are applying more holistic approaches to the feedback control of the air transportation system [12-13]. Many of these approaches rely on economic feedback, including the cost of delays to the airlines. Therefore, understanding the true cost of a delay is not only of interest

to the airlines that incur these costs but is essential for air transportation management.

We begin this study by considering only the direct costs to the airlines of such delays. We then apply estimates of network effects on delay costs based on a study performed by American Airlines [6]. Future work will examine the social costs of such delays, i.e. the resulting economics costs to the various regions and other industries.

In general a flight can be delayed due to several reasons, mainly:

- Mechanical problems with the aircraft.
- Schedule disruption due to bad weather or air traffic management initiatives (Ground Delay Programs (GDPs) or Air Flow Programs (AFPs)).
- Misaligned crew/ aircraft due to previous delayed flight

Weather is a major cause of delay as it reduces the capacity of both the airspace and the runways. At several highly utilized airports, over-scheduling also plays role in causing delays. Based on weather forecasts and schedules, air traffic management estimates the resulting reduction in capacity within various segments of the airspace and at a variety of airports. It announces Ground Delay Programs (GDPs) that hold aircraft at the departing airport, in order to have the flying aircraft better match the capacity of the system. For capacity reduction in air, Air Flow Programs (AFPs) are employed that suggest/announce alternative routes for the flights. Holding at a gate is both cheaper and safer than airborne holds, and allows the system to be better managed. Finally, the delays already described induce future delays in the system, because the aircraft or crews may not arrive at their next assignment on time. Even when the crew does arrive, they may not be able to work another flight because they have exceeded their allowable working hours.

We base our work on a final report evaluating true cost of flight delays that was prepared by the Performance Review Unit, EuroControl in 2004[2]. This EC report

This research was partially funded under NASA grant NNX09AB20A and NASA grant NNX07AT23A.

describes a methodology and presents results detailing the cost to airlines of delays during various segments of a trip. The costs are divided into short delays (less than 15 minute) and long delays (greater than 65). The report provides the resultant multiplier (Euros per minute) for any such segment. The types of delays considered include gate delay, access to runway delay (both taxi in and out delays), on routes delays, and landing delays (circling or longer flight paths to overcome congestion while approaching the airport). The data used in the study consisted of data collected from European airlines, air traffic management as well as interviews and surveys conducted by the research team. However, their model is not explicitly defined and therefore no sensitivity analysis is possible to changes in the cost factors (e.g. fuel). Furthermore, the model is generated based on data from EU airlines and is stated in terms of costs in 2003 Euros.

The motivation of this paper is therefore to:

- Better understand each of the cost factors involved.
- Develop a model that includes each of the cost factors
- Make the model consistent to US data.

This paper is organized as follows. Section II describes the EC report, Section III provides our methodology for determining the cost components and multipliers that make up the final multipliers used in the EuroControl report and describe our validation of the new model on European data from the period of the EC report. In Section IV and V, we apply our methodology to 8 weather days at Philadelphia as a case study and show the resulting delay costs for these flights. Section VI provides conclusions and Section VII points out the future research.

II. EUROCONTROL PERFORMANCE REVIEW UNIT REPORT (EC REPORT)

The EC report specifies that delays incurred can be of two types: *tactical delay* and *strategic delay*. The report makes the distinction between tactical delays (delays encountered that are greater than the announced schedule, i.e. delays above the anticipated padding of the schedule) and strategic delays (i.e. the delay relative to an unpadded schedule). Both US and European airlines increase the arrival time over unimpeded time so that they can report “on time” performance even when the system is over-capacitated. Another distinction that the report makes is between *gate-to-gate (or single flight) delays* and *network-level delays*. The gate-to-gate delay is the delay that an individual flight incurs based on the environment it encounters, while the network delays are the effects that the flight causes to the rest of the network. The cost of delay discussed in the EC report is the tactical primary delay. In the report, two types of delays have been chosen for demonstration: delays of *short* duration (15 minutes or less) and delays of *long* duration (65 minutes or more). Similarly three cost scenarios have been used to “allow more realistic ranges of values”.

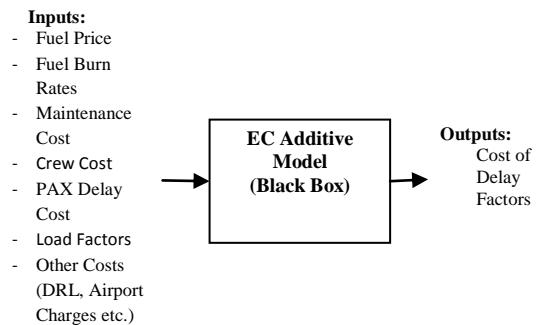


Figure 1: EuroControl (EC) Model

TABLE 1: LOW, BASE AND HIGH COST SCENARIOS (FROM TABLE 2-5 OF [2])

Factor	'short' delay type: '15 minutes' basis			'long' delay type: '65 minutes' basis		
	low	base	high	low	base	high
load factor	50%	70%	90%	50%	70%	90%
transfer passengers	15%	25%	35%	15%	25%	35%
arrival / departure ^(a)	domestic	EU	non-EU	domestic	EU	non-EU
turnaround time ^(a)	60 mins	60 mins	60 mins	60 mins	60 mins	60 mins
parking ^(g)	remote	pier	pier	remote	pier	pier
fuel price ^(c)	low	base	high	low	base	high
weight payload factor	50%	65%	80%	50%	65%	80%
airborne fuel penalty ^(f)	none	none	applied	none	none	applied
handling agent penalty	none	none	none	none	none	charged
extra crew costs ^(d)	none	none	low	none	medium	high
airport charges	averaged	averaged	max/2	averaged	averaged	max/2
pax cost of delay to AO, EUR/min ⁽ⁱ⁾	0	0	0.05	0.32	0.40	0.48
aircraft depreciation, rentals & leases ⁽ⁱ⁾	Strategic cost model used: please see Annex O			Strategic cost model used: please see Annex O		
BHDOC ^(b) scenario	low	base	high	low	base	high
maintenance ^{(e) (h)}	15%	15%	15%	15%	15%	15%

The EC report describes the model as an additive model where each component describes some proportions of the total cost. Table 1 shows what costs factors are included as inputs in these cost scenarios under different delay characteristics. For details, see [2]. Figure 1 details the inputs and outputs of their model.

Further exploring their cost factors reveals the following costs involved:

- **Fuel cost:** The report provides different fuel burn rates for each aircraft type studied and for all segments of the flights. The prices for all cost scenarios and conversion rate from Euro to Dollars are also provided. (See Table 2-12 and Annex C in [2]).
- **Extra Crew cost:** The report defines extra crew cost as extra cost paid in addition to the usual flight and cabin crew salaries and expenses. It may include employing

additional crew (both flight and cabin crew) or incurring additional pay for regular crews due to unexpected increases in hours worked. The report does not specify exactly the methodologies used to obtain the crew cost component of the multiplier in order to preserve confidentiality of airline data. However, the report describes under what circumstances the cost factors will be increased (refer to Table 1 of this paper).

- Maintenance cost:** The maintenance cost is defined to be the cost of maintaining both the airframe and power plant of the aircraft. The additional maintenance cost incurred for a one-minute delay is stated in the report as approximately 15% of the Block Hour Direct Operating Cost (BHDOC). The proportions of how maintenance cost is divided into different segments of the flights are given in Annex J of [2]. BHDOC's are given in the report for *low*, *base* and *high* cost scenarios for the 12 different aircraft systems studied (see Table 2-11 in [2]).
- Depreciation Cost:** The report assumes that there is no additional depreciation cost caused by delays. Thus, the depreciation component of total delay is taken to be zero for all segments and cost scenarios.
- Passenger Delay Cost:** Passenger Delay cost (or PAX delay cost) is defined as the compensation paid by the airlines to passengers who have experienced delayed flights. Passenger Delay (in cost per passenger per minute) is given as: none for *low* and *base* cost scenarios, 0.05 for the *high* cost scenario for 15 minutes of delay and 0.32, 0.40 and 0.48 for *low*, *base* and *high* cost scenarios respectively for 65 minutes delay. The load factors assumed are: 50% for *low*, 70% for *base* and 90% for *high* cost scenarios.
- Other Costs:** This factor is a catch-all component that attempts to include any other cost factors mentioned in Table 1 (such as parking, airport charges, handling agent penalty, weight payload factor etc.). No specific cost factors were given in the report, except details for different Airport charges at different EU airports are provided (see Annex L in [2]).

TABLE 2: TACTICAL GROUND DELAY COSTS: AT-GATE ONLY
(WITHOUT NETWORK EFFECTS)

Aircraft and number of seats		based on 15 minutes' delay			based on 65 minutes' delay		
		cost scenario			cost scenario		
		low	base	high	low	base	high
B737-300	125	0.6	0.9	14.5	20.4	44.6	82.8
B737-400	143	0.6	0.9	15.8	23.7	50.3	92.3
B737-500	100	0.6	0.8	13.8	16.6	38.2	73.5
B737-800	174	0.5	0.8	17.1	28.4	58.6	105.2
B757-200	218	0.6	1.0	20.2	35.6	71.7	126.0
B767-300ER	240	0.6	1.2	27.8	39.2	84.9	155.1
B747-400	406	1.8	2.2	49.0	67.1	142.2	258.7
A319	126	0.6	0.9	14.7	20.8	45.0	83.8
A320	155	0.6	0.9	16.3	25.3	53.5	96.5
A321	166	0.7	1.0	16.6	27.3	56.3	100.7
ATR42	46	0.4	0.6	8.6	7.8	19.7	40.6
ATR72	64	0.5	0.6	9.6	10.7	25.0	48.6

Based on the analysis done, the EC report provides cost of delay factors (in Euros). The delay is divided into three segments of the flight; delay on the ground at the gate (Table 2), delay while taxiing at either airport (Table 3) or delay while airborne (en-route and holding, Table 4). These segments were chosen for discussion because they reflect the fidelity of publically available data.

TABLE 3: TACTICAL GROUND DELAY COSTS: TAXI-ONLY
(WITHOUT NETWORK EFFECTS)

Aircraft and number of seats		based on 15 minutes' delay			based on 65 minutes' delay		
		cost scenario			cost scenario		
		low	base	high	low	base	high
B737-300	125	3.0	4.6	19.0	22.9	48.4	87.1
B737-400	143	3.0	4.7	20.3	26.1	54.1	96.6
B737-500	100	3.0	4.6	18.2	19.0	42.0	77.8
B737-800	174	2.9	4.5	21.6	30.8	62.3	109.5
B757-200	218	3.4	5.3	24.9	38.4	76.0	131.0
B767-300ER	240	4.5	7.2	34.0	43.2	91.0	162.1
B747-400	406	10.6	15.9	61.7	76.4	156.3	276.2
A319	126	2.6	4.1	18.4	22.8	48.2	87.4
A320	155	2.6	4.0	20.1	27.3	56.7	100.1
A321	166	3.0	4.7	20.9	29.7	60.1	105.0
ATR42	46	0.6	0.9	8.2	7.9	20.0	40.0
ATR72	64	1.1	1.8	10.3	11.4	26.1	49.2

TABLE 4: TACTICAL AIRBORNE DELAY COSTS AND HOLDING
(WITHOUT NETWORK EFFECTS)

Aircraft and number of seats		based on 15 minutes' delay			based on 65 minutes' delay		
		cost scenario			cost scenario		
		low	base	high	low	base	high
B737-300	125	9.5	14.8	34.1	28.9	57.8	102.3
B737-400	143	9.2	14.3	34.6	32.0	63.3	111.4
B737-500	100	8.9	13.7	31.6	24.5	50.3	91.1
B737-800	174	7.8	12.5	33.1	36.5	71.3	122.6
B757-200	218	10.3	16.1	40.7	46.2	88.2	149.7
B767-300ER	240	14.2	22.5	57.1	54.2	108.4	189.5
B747-400	406	27.6	42.2	102.4	97.5	188.8	332.7
A319	126	7.1	11.1	29.1	28.1	56.4	101.3
A320	155	7.7	12.0	32.3	32.9	65.3	115.0
A321	166	9.5	14.9	36.2	36.5	70.7	122.2
ATR42	46	1.6	2.6	10.8	9.1	21.9	42.8
ATR72	64	2.2	3.4	12.8	12.7	28.1	52.6

Since the data is in Euros, we have used the conversion rate of 1 Euros = 1\$ (as used by the report).

One point worth mentioning is that the findings of the report are for EU airports only. We validate their cost factors by applying the imputed cost factors to their data. However, once we have obtained these costs factors, when applying the formulas to US data, we recognize the differences between the US and European system and adjust the calculations accordingly to reflect these differences. For example, passenger compensation costs incurred to the airline in US are far lower than that of EU (due to EU Passenger Bill of Rights or PBR). Similarly, aircraft spend more time taxiing out in the US than in Europe. Also, in the

US, Air Traffic Management imposes greater ground delay programs in order to assure that there is little circling at the destination airport. The EC report specifically comments on this difference noting that, on average, the amount of en route delay is greater than the amount of ground delay for European flights.

III. METHODOLOGY

A. Regenerating the EC Model

For our analysis, we start with a similar additive general model for each of the different segments paired with the different cost scenarios that include all the different cost factors. Due to the fidelity of the available US data, we divide the flight into three segments; gate, taxi and en-route (which includes both airborne and holding). For each of these segment, three cost scenarios and two range delays are provided, hence for each of these 18 different cases (segments x cost scenarios x delay ranges), we have the following model:

$$\begin{aligned} C_{\text{delay}} = & c_{\text{fuel}} \times \text{fuel burn rate} \times \text{fuel price} \\ & + c_{\text{crew}} \times \text{crew cost} \\ & + c_{\text{maintenance}} \times \text{maintenance cost} \\ & + c_{\text{other}} \times \text{other cost} \\ & + c_{\text{pax}} \times \text{PAX delay cost} \times (\# \text{ seats}) \times \text{load factor} \end{aligned}$$

All costs factors are in minutes. The coefficients in this cost model were determined so that we obtain a good fit to the EC data, as presented in the report. The validation was done using each of the three scenarios (low, base and high) and each of the 12 aircraft types in that report. Since fuel burn is directly applied in the formulation with no multiplier, the fuel coefficient (i.e. c_{fuel}) is 1 for airborne and taxi segments and 0 for gate segment. We fix the catch-all category "Other Costs" to be \$1.6¹ and the other cost coefficient (i.e. c_{other}) to be 0.15 for gate segment and 0 otherwise, since these are consistent with the EC report. The PAX cost coefficient (i.e. c_{pax}) is set to be 1 for validation purposes. However, we revise this when applying it to US data. Hence, the only two variables that we need to determine are the coefficients for crew costs and for maintenance cost.

Specifically, we need to determine the factors for all combinations of the two delay ranges, the three scenarios, and the three flight segments, or 18 (possibly different) sets of coefficients in all. We note, however, that we have assumed that the coefficients were independent of aircraft type.

B. Modify Model for US Data

In order to apply this model to the US data, we made the following changes that are more consistent to the US airlines.

- We used cost factors from the BTS P52 database (fuel price, crew and maintenance cost) [3].
- We used the fuel burn rate while en route from the BTS P52 database. For taxi burn rate, we used ICAO engine emissions databank. (See [5]).
- We set the PAX delay cost coefficient to 0, since in US; it is not incurred by the airlines.

For other delay ranges, we used the following formulas: for any delay less than or equal to 15 minutes, we used 15 minutes cost factor, similarly for any delay above 65 minutes, we used the cost factor for 65 minutes and above delay. For delays between 15 and 65 minutes, we interpolate using the two data points.

For the network effect of these delays, we use the delay multipliers based on American Airlines case study (see Table 2-20 in [2] or [6]).

C. Case Study

Finally, as a case study, we applied our cost factors to 8 representative weather days at Philadelphia Airport (PHL) that have cancellation rates ranging as low as 1% to a very bad day where 68% of the flights were cancelled. The data is taken from ASPM database [7]. We used 2/13/2007, 3/16/2007, 3/23/2007, 8/9/2007, 2/1/2008, 2/12/2008, 2/22/2008, and 6/23/2008 for the case study. We chose these days because, in every case, there were Ground Delay Programs that forced large delays.

Our next section describes all the results and observations we found during our analysis.

IV. RESULTS

Before beginning the work to determine the cost coefficients for the new model, we first examined whether overall cost factors in the US appear to be similar to those incurred in Europe. We computed, based on the EC factors, the different types of delay cost (gate, taxi and airborne-and-holding) for the given 12 aircrafts and compared it with the average operational cost per minute using P52 [3] data from the BTS database for US airlines.

¹ This represents the other cost of operations which is \$1.87 in 2008 Dollars(see [4])

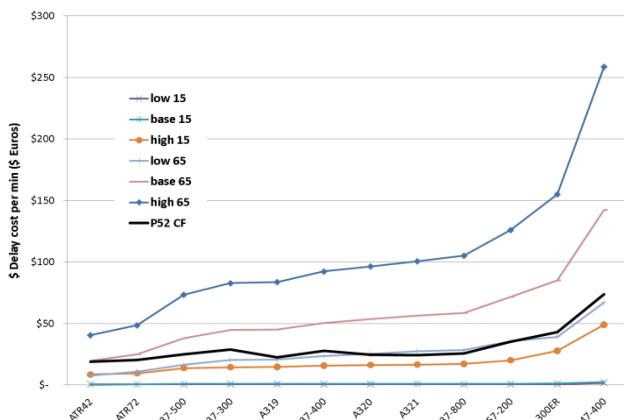


Figure 2: Tactical Ground Delay costs: gate only (without network effect) vs. Operational costs

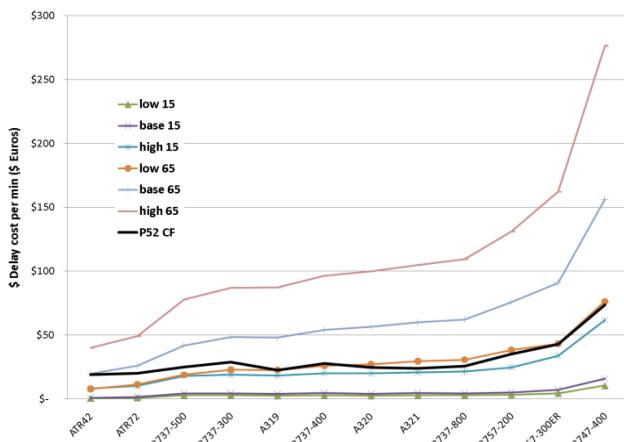


Figure 3: Tactical Ground Delay Costs: Taxi only (without network effect) vs. Operational costs

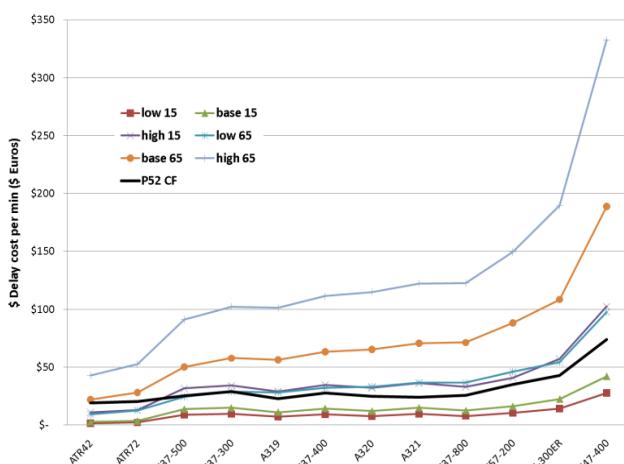


Figure 4: Tactical Airborne Delay Costs en-route and holding (without network effect) vs. Operational costs

cost factors are consistent with the operational costs in the US.

Next we worked to determine the multipliers for crew and maintenance costs that would, when combined with the other factors sum to the resultant multipliers provided in the EC report. Table 5 provides the computed multipliers. To illustrate how close we come to the multipliers provided in the report, we combine the individual multipliers into the summarized single multiplier for total delay cost and compare this multiplier to that provided in the EU report. These resultant multipliers are provided in Tables 6-8 below. Green cells indicate the cases where EC cost factors are 10 % higher than ours; Red cells indicate the cases where our cost factor is 10% higher than EC reports. All the remaining cells have values with difference of within 10%. There are instances where the variations are off by more than 10%, but mostly they are in the 15 minute delay category and mostly, our numbers are lower than those of the EC estimates. We assert, therefore, that the derived numbers are likely to estimate well the costs of long delays.

TABLE 5: COEFFICIENTS COMPUTED ON FITTING THE EC DATA

Gate Only						
Cost Factors	Based on 15 Minutes Delay			Based on 65 Minutes Delay		
	cost scenario			cost scenario		
	Low	Base	High	Low	Base	High
Fuel	0	0	0	0	0	0
Crew	0	0	0.5	0	0.85	2
Maintenance	0.02	0.02	0.05	0.05	0.05	0.05
PAX delay	1	1	1	1	1	1
Other	0.15	0.15	0.15	0.15	0.15	0.15
Taxi Only						
Cost Factors	Based on 15 Minutes Delay			Based on 65 Minutes Delay		
	cost scenario			cost scenario		
	Low	Base	High	Low	Base	High
Fuel	1	1	1	1	1	1
Crew	0	0	0.5	0	0.85	2
Maintenance	0.02	0.02	0.05	0.05	0.05	0.05
PAX delay	1	1	1	1	1	1
Other	0	0	0	0	0	0
En-route						
Cost Factors	Based on 15 Minutes Delay			Based on 65 Minutes Delay		
	cost scenario			cost scenario		
	Low	Base	High	Low	Base	High
Fuel	1	1	1	1	1	1
Crew	0	0	0.5	0	0.85	2
Maintenance	0.02	0.02	0.1	0.05	0.05	0.1
PAX delay	1	1	1	1	1	1
Other	0	0	0	0	0	0

TABLE 6: TACTICAL GROUND DELAY COSTS: GROUND ONLY. DIFFERENCE BETWEEN EC AND OUR COST FACTORS FOR GIVEN 12 AIRCRAFTS
(COMPARED TO TABLE 2 OF THIS PAPER)

Aircraft and Number of seats		Based on 15 min. delay			Based on 65 min. delay		
		cost scenario			cost scenario		
		low	base	high	low	base	high
ATR42	46	0.30	0.31	0.20	(0.02)	0.08	0.12
ATR72	64	0.05	0.14	0.07	(0.02)	0.02	0.03
B737-500	100	(0.05)	(0.01)	0.03	(0.03)	0.01	0.03
B737-300	125	(0.05)	(0.01)	(0.02)	(0.03)	(0.02)	(0.03)
A319	126	0.03	0.09	(0.01)	(0.02)	(0.02)	(0.03)
B737-400	143	(0.05)	0.01	(0.01)	(0.02)	(0.02)	(0.02)
A320	155	0.03	0.07	(0.01)	(0.02)	0.00	(0.04)
A321	166	0.02	0.08	(0.04)	(0.01)	(0.03)	(0.05)
B737-800	174	(0.09)	(0.03)	0.01	(0.02)	0.02	0.00
B757-200	218	(0.03)	0.03	(0.03)	(0.01)	(0.02)	(0.03)
B767-300E	240	(0.09)	(0.00)	0.00	(0.02)	0.02	(0.02)
B747-400	406	(0.12)	(0.10)	0.08	(0.03)	0.03	0.08

TABLE 7: TACTICAL GROUND DELAY COSTS: TAXI ONLY. DIFFERENCE BETWEEN EC AND OUR COST FACTORS FOR GIVEN 12 AIRCRAFTS
(COMPARED TO TABLE 3 OF THIS PAPER)

Aircraft and Number of seats		Based on 15 min. delay			Based on 65 min. delay		
		cost scenario			cost scenario		
		low	base	high	low	base	high
ATR42	46	0.30	0.31	0.20	(0.02)	0.08	0.12
ATR72	64	0.05	0.14	0.07	(0.02)	0.02	0.03
B737-500	100	(0.05)	(0.01)	0.03	(0.03)	0.01	0.03
B737-300	125	(0.05)	(0.01)	(0.02)	(0.03)	(0.02)	(0.03)
A319	126	0.03	0.09	(0.01)	(0.02)	(0.02)	(0.03)
B737-400	143	(0.05)	0.01	(0.01)	(0.02)	(0.02)	(0.02)
A320	155	0.03	0.07	(0.01)	(0.02)	0.00	(0.04)
A321	166	0.02	0.08	(0.04)	(0.01)	(0.03)	(0.05)
B737-800	174	(0.09)	(0.03)	0.01	(0.02)	0.02	0.00
B757-200	218	(0.03)	0.03	(0.03)	(0.01)	(0.02)	(0.03)
B767-300E	240	(0.09)	(0.00)	0.00	(0.02)	0.02	(0.02)
B747-400	406	(0.12)	(0.10)	0.08	(0.03)	0.03	0.08

TABLE 8: TACTICAL AIRBORNE DELAY: ENROUTE AND HOLDING.
DIFFERENCE BETWEEN EC AND OUR COST FACTORS FOR GIVEN 12 AIRCRAFTS
(COMPARED TO TABLE 4 OF THIS PAPER)

Aircraft and Number of seats		Based on 15 min. delay			Based on 65 min. delay		
		cost scenario			cost scenario		
		low	base	high	low	base	high
ATR42	46	0.08	0.07	0.15	0.00	0.07	0.11
ATR72	64	0.00	(0.02)	0.04	(0.00)	0.02	0.04
B737-500	100	0.15	0.14	0.12	0.02	0.03	0.05
B737-300	125	0.13	0.13	0.08	0.01	(0.00)	0.00
A319	126	(0.10)	(0.11)	(0.06)	(0.01)	(0.03)	(0.02)
B737-400	143	0.11	0.10	0.06	0.01	(0.00)	0.00
A320	155	(0.04)	(0.04)	(0.02)	(0.01)	(0.00)	(0.02)
A321	166	0.01	0.01	(0.02)	(0.00)	(0.03)	(0.03)
B737-800	174	(0.12)	(0.09)	(0.04)	(0.01)	0.01	(0.00)
B757-200	218	(0.09)	(0.09)	(0.07)	(0.01)	(0.03)	(0.03)
B767-300E	240	(0.11)	(0.11)	(0.05)	(0.01)	0.01	(0.02)
B747-400	406	(0.20)	(0.22)	(0.03)	(0.02)	0.01	0.08

V. APPLICATION TO US DATA

When using the same model but using fuel burn rates as reported in US databases, we observed that fuel burn rates reported in the US are lower than reported in the EC report.

This means that even using the model postulated in the EC report, we will have slightly lower costs for equivalent delays than that of the EC report. Table 9 shows the final cost factors computed using the model with our data. We have used the coefficients for the base cost scenario.

We next apply these cost factors to the 8 weather days at PHL. We first compute the non-network costs and then, use the delay multipliers from American Airlines case study (Table 2-20 in [2] or [6]) to compute the network delays and their resulting costs. Figures 5-9 provide some of the results of this case study.

TABLE 9: OUR COEFFICIENTS FOR DIFFERENT COST FACTORS FOR US DATA

Cost Factor	Gate		Taxi		En-route	
	15 min	65 min	15 min	65 Min	15 min	65 min
Fuel	0	0	1	1	1	1
Crew	0	0.85	0	0.85	0	0.85
Maintenance	0.02	0.05	0.02	0.05	0.02	0.05
PAX	0	0	0	0	0	0
Other	0.15	0.15	0	0	0	0

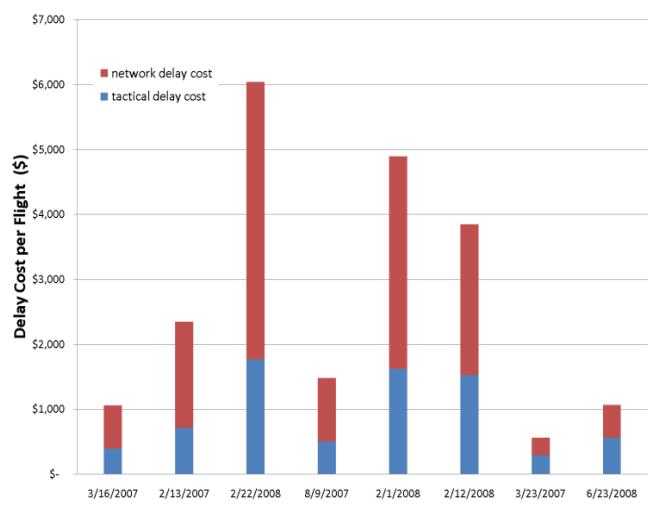


Figure 5: Cost of delay per Flight for observed days

Looking at the cost of delay for each observed day (Figure 5), we see that the cost of delay is not proportional to the proportion of flights cancelled that day. For example, day “2/22/2008”, despite having only 22% cancelled flight has the highest cost of delay while day “3/16/2007” with the highest number of cancelled flight has very low cost of delay. One possible explanation for this result is that all cancelled flights are recorded as having zero delay. Thus, a

day with more delays but lower cancellations will have lower costs. Future research will evaluate how to better cost out cancelled flights.

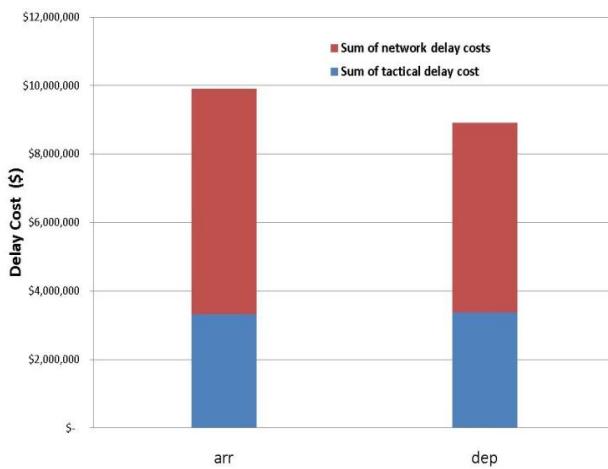


Figure 6: Delay costs (arrivals vs. departures at PHL)

The total costs of delay for departures and arrivals at PHL are very similar, Figure 6. However, arrivals show more network delay costs.

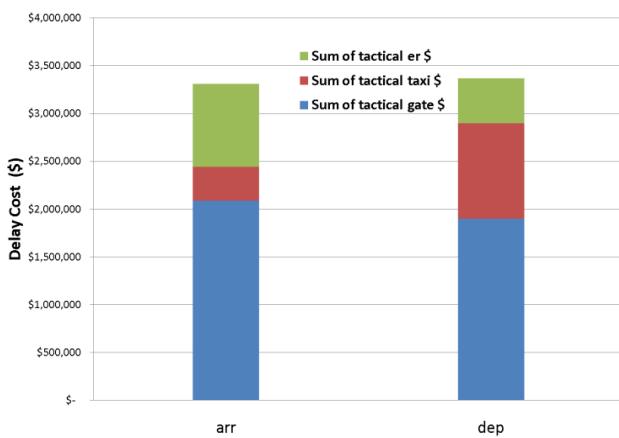


Figure 7: Arrival vs. Departure Tactical Delay costs across all segments of flight

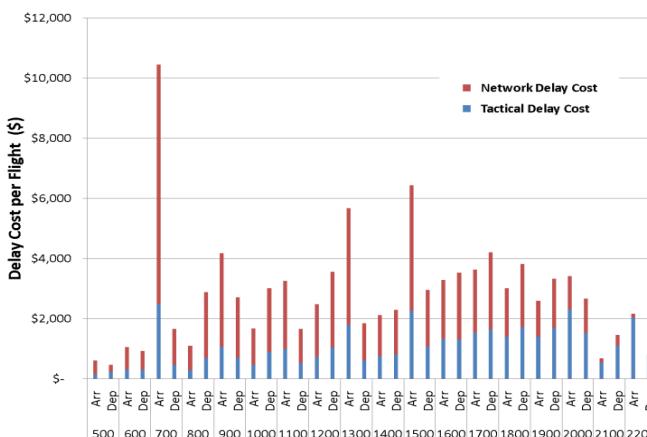


Figure 8: Departure vs. Arrival cost of delay per flight by time of day

Comparing the total cost of primary delay for arrivals vs. departures at PHL for the segments of flights, Figure 7, we see the total delay cost is approximately the same. However, arrivals show slightly more gate delay costs and significantly more airborne delay costs than are observed with departures. And departures show significantly more taxi delay costs than are observed with arrivals.

Analysis of the departure and arrival delay costs per flight by time of day is shown in Figure 8. Arrival delay costs per flight are shown to be much higher for 0700, 1300 and 1500 hrs arrivals.

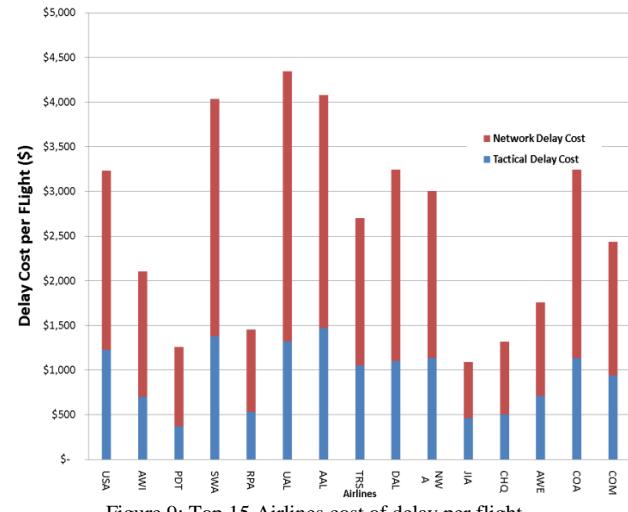


Figure 9: Top 15 Airlines cost of delay per flight

Analysis of the top 15 airlines cost of delay per flight is shown in Figure 9. One interesting result shows that not all airlines incur similar delay costs at PHL. Southwest, United Airlines, Delta Airlines and American Airlines all have higher costs of delay at PHL than does the dominant carrier, US Air. Also, the regional airlines have lower costs of delay than the larger ones.

VI. CONCLUSIONS

From our analysis, we conclude the following:

- The cost factors from the EC report and costs as reported by US carriers in BTS P52 database follow similar trends. Thus, the general approach taken by [3] the EC report can be applied, with minor modifications, to compute the cost of delays for US flights
- We determined appropriate multipliers for crew and maintenance costs that, when combined with the other factors produced multipliers close to those reported in the EC report.
- The US data shows that very long taxi delays at PHL, which has one dominant airline, US Air. We presume that this airline schedules its flights at peak times in order to restrict competition.
- The cost of delay is not proportional to the flights flown. One reason for this non-intuitive result is that when a flight is cancelled, it is recorded as having zero delay.

Future research will address how to cost cancelled flights.

- We observe peaking at PHL and this scheduling of departures above runway capacity results in larger delay costs. The network delays are not necessarily larger for these peak times.
- One interesting result shows that not all airlines incur similar delay costs at PHL. Southwest, United Airlines, Delta Airlines and American Airlines all have higher costs of delay at PHL than does the dominant carrier, US Air. Also, the regional airlines have lower costs of delay than the larger ones. Here too, the issue may be one of the way in which the data is recorded. The regional jets are more likely to be cancelled than the larger aircraft and, when cancelled, the data records such flights as having zero delay.
- Our calculations of the cost of delayed flights (but not cancelled flights) total \$18M for these 8 days.

Many economic modeling and analysis efforts require a good understanding of the costs that an airline will incur when it experiences delays at the gate, while taxiing or while en-route. This paper has presented a relatively straightforward mechanism for calculating such costs and for predicting how such costs are likely to increase when there is a change in fuel costs, aircraft type, or other major alternative in the cost structure. It is informative in explaining why airlines are currently down-gauging the size of the aircraft used even at airports with substantial capacity restrictions.

VII. FUTURE WORK

We intend to both expand and apply this model in a variety of efforts currently underway:

- We need to devise a mechanism for including the costs of cancellations in the overall cost calculations. The research of Hansen et al. [9], Wang, et al. [10] and Barnhart and Batu [11] will assist in this effort.
- We wish to apply the model and investigate its sensitivity to significant cost changes in fuel or crew, and changes in aircraft usage. By separating the cost factors into their component parts, we are now able to apply the model to aircraft types not studied in the EC model. For application to the US environment, this capability is imperative.
- We will next apply this model to a variety of different airports and see how airline costs vary based on different mixes of aircraft, varying amounts of airline dominance, and alternative government policies (such as slot-controls, rules about entry into the airport, etc.)
- We intend to examine if, based on these costs, we can predict which flights are most likely to be cancelled or delayed when weather conditions result in the initiation of a Ground Delay Program.
- Once this model has been validated for a variety of different congestion scenarios and airports, we intend to

include the model as part of a larger equilibrium model that predicts the actions of airlines under various policy decisions. See [8] for more on this effort.

- We intend to use this as a tool in a congestion-pricing model to determine the flights that are most likely to be cancelled first when capacity at an airport is reduced, and thereby to determine the prices that would be needed to have supply approximately equal demand if congestion pricing were imposed at some airport imposed.

ACKNOWLEDGMENT

We gratefully acknowledge the support and assistance that Michael Bloem, Rosa Osegura-Lohr and Mike Madson have provided throughout this research effort. We thank George Hunter and Huina Gao, both of Sensis Inc, for bringing the EuroControl research and American Airlines study to our attention.

REFERENCES

- [1] C. E. Schumer, "Flight Delays Cost Passengers, Airlines and the U.S. Economy Billions". A Report by the Joint Committee Majority Staff, May 2008.
- [2] Performance Review Unit, Eurocontrol, "Evaluating the True Cost to Airlines of One Minute of Airborne or Ground Delay," University of Westminster Final Report, May, 2004.
- [3] (Online) Bureau of Transportation Statistics (BTS) Databases and Statistics. <http://www.transtats.bts.gov/>
- [4] (Online) Air Transport Association of America, Inc (ATA), cost of delays. <http://www.airlines.org/economics/cost+of+delays/> (2008).
- [5] (Online) ICAO Engine Emissions databank, ICAO Committee on Aviation Environmental Protection (CAEP), hosted on UK Civil Aviation Authority, <http://www.caa.co.uk/default.aspx?catid=702> (Updated Feb 2009).
- [6] Beatty R, Hsu R, Berry L & Rome J, "Preliminary Evaluation of Flight Delay Propagation through an Airline Schedule", 2nd USA/Europe Air Traffic Management R&D Seminar, December 1998
- [7] (Online) Aviation System Performance Metrics (ASPM)-Complete. FAA, <http://aspm.faa.gov/aspm/entryASPM.asp>
- [8] Gao, Huina, Hunter, George, Barardino, Frank and Hoffman, Karla (2010) "Development and Evaluation of Market-Based Traffic Flow Management Concepts" Technical report submitted to the 10th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference.
- [9] Hansen, Mark M., David Gillen and Reza Djafarian-Tehrani (2001) "Aviation infrastructure performance and airline cost: A statistical cost estimation approach" Transportation Research Part E: Logistics and Transportation Review 37(1) 1-23.
- [10] Wang, D., L. Sherry and G. Donohue. Passenger Trip Time Metric for Air Transportation. The 2nd International Conference on Research in Air Transportation, Belgrade, Serbia and Montenegro, June 2006.
- [11] Bratu S. and C. Barnhart (2005), "An analysis of passenger delays using flight operations and passenger booking data", Air Traffic Control Quarterly, 13.
- [12] G. L. Donohue, R. D. Shaver III, "Terminal Chaos: Why U.S. Air Travel is Broken and How to Fix It", American Institute of Aeronautics & Astronautics, Library of Flight, Editor: Ned Allen Spring, 2008.
- [13] M. O. Ball, L. M. Ausubel, F. Berardino, P. Cramton, G. Donohue, M. Hansen, K. L. Hoffman, "Market-Based Alternatives for Managing Congestion at New York's LaGuardia Airport, Optimal Use of Scarce Airport Capacity", Proceedings of AirNet Annual Conference, The Hague, April 2007.

Track 3

Airport Design and Operations

Airport Ground Access and Egress Passenger Flow Model (AGAP)

Dr. Milan Stefanik, Dr. Benedikt Badanik, Martin Matas
 Air Transport Department
 University of Zilina
 Zilina, Slovakia
milan.stefanik@fpedas.uniza.sk

Abstract— The problems of airport landside capacity assessment are of industry-wide interest. Evaluation of landside capacity enables airport operators and airport designers to identify passenger and baggage flow bottlenecks, identify the primary cause of bottlenecks formation and take measures mitigating the impact of bottlenecks on the airport terminal operation.

Many studies dealing with the problems of airport landside capacity are focused mainly on the processing part of the airport terminal and consider the airport terminal to be an isolated system. Even the most of models of airport landside operations developed using either generic or dedicated simulation software packages (e.g. PaxSim, SLAM, WITNESS, ARENA or EXTEND) are designed for simulating the passenger and baggage flows only between curb-side and apron. Although this approach provides valuable data concerning capacity, delays or processing bottlenecks, in some cases identified capacity constraints are only the symptoms of the actual problem. In order to discover the cause of the problem, it is necessary to consider the airport terminal as an integral part of much more complex regional, national or international transportation system.

This article reflects the above mentioned requirements and introduces an innovative approach to passenger and baggage flow simulation based on the fact that airport terminal is considered as an integral part of air passenger door-to-door transportation process.

Keywords-airport ground access; fast-time simulations; airport capacity enhancement; door-to-door transportation process.

I. INTRODUCTION

The air transport in Europe as well as worldwide has been undergoing a rapid and continuous growth in the recent years and it is anticipated that by 2030 there will be between 1.7 and 2.2 times the number of flights in Europe seen in 2007 [1]. One of the most serious problems of air traffic system that will have to be solved in the following years is the capacity issue, and that applies to both airports and airspace. The airports are generally considered as a principal constraint to traffic growth and increasing demand will definitely lead to congestion of airports and Terminal Manoeuvring Areas (TMAs) and to generation of delays. It is expected that despite planned airport infrastructure investments, in 2030, 19 European airports will be operating at full capacity eight hours a day, every day of the year and involving 50% of all flights each day. If the most challenging scenario is considered, there will be as much as 39

airports in Europe operating at their full capacity and involving as much as 70% of all flights [2].

However, this trend does not necessarily mean that duplication infrastructure will be required to accommodate the demand in 2030. Implementation of measures that lead to more efficient traffic flows and better utilisation of existing infrastructure (ACE, CDM, TAM etc.) seems to be the right approach for solving the current and future capacity issues. In fact, thanks to these measures the efficiency with which the physical infrastructure at airports is used is increasing significantly. Thus despite the absence of obvious investments such as new runways or terminal buildings the Europe's most congested airports keep their ability to accommodate the growing demand. Needless to point out that these airports have been considered as saturated for years [3].

However, the airside capacity is not the only problem the European airports currently face to. After September the 11th and after security alerts in UK during summer 2006, the airport security became a priority and it has affected passenger flows within the airport terminals. The security procedures that were introduced at European airports after summer 2006 caused the 35% dwell time increase [4]. However, the long queues at check-in counters and at security checkpoints are not the only issues the airport operators have to deal with. A large percentage of private vehicles access trips at many airports lead to congestion of airport access roads and car parks. Moreover, high share of individual car access trips has negative impact on the environment. At many airports, the ground access trips of private cars associated with the airport operation generate a greater share of air pollution than the aircraft movements [5].

In order to increase the capacity and thus keep the capability to accommodate the growing demand, 138 European airports reported that they are planning significant investments. If these plans can be delivered, these 138 airports in total projected that their capacity would be 41% higher in 2030 compared to 2007. These plans include investments in building new runways and in improving airside (taxiways, aprons etc.) and landside (passenger terminals etc.) infrastructure [2].

Despite the planned investments into airport infrastructure, the airports will become the principal bottleneck of the air transport network that will generate enormous delays and unaccommodated flight demand. It is anticipated that in 2030, the highly-congested air traffic network will generate 2.3

million of unaccommodated flights, which will be approximately 11% of the overall flight demand [2].

However, the headless investments into airport infrastructure do not seem to be the right solution of the capacity problems due to time and geographical flight demand imbalance. It is necessary to realize that the air transport is the subject of significant seasonal, daily and hourly demand fluctuations. It means that many airports are congested during traffic peaks but fairly deserted during times that are not so attractive for passengers. In other words, there are times of day when the traffic is very high and reaches critical hourly values for either the airside or landside (or both); these are called peak hours. Nevertheless, looking at the annual operation many airports can be far from hitting the line. The peak hours simply reveal the bottlenecks of airports. Moreover, thanks to the existence of geographical flight demand imbalance only top 133 out of more than 2000 European airports carry as much as 90% of the ECAC IFR traffic [6].

It means that there is a big mismatch between when and where the capacity is available, and when and where the demand is present. This leads to inefficient utilisation of the existing airport infrastructure. Taking this into account, the following methods have been identified as measures that could be potentially used for mitigating the effects of the congested European air transport network [2]:

- Schedule smoothing: Move flights to times of the day when more capacity is available.
- Alternative airports: Move excess traffic either to secondary or to regional airports.
- Larger aircraft: Use larger aircraft to reduce daily frequencies on congested airports.
- Investments into high-speed train networks: Replace busy, short-range airport pairs flights by high-speed train connections.
- Exploitation of benefits of SESAR: The SESAR programme will be making a major contribution to the efficiency of air traffic management in the 2020 – 2030 timeframe.

All the above listed methods consider air transport network as an isolated and independent transportation system and the problems of airport terminals and airport ground access are being underestimated. However, it is necessary to realize that due to physical and nuisance constraints the airports have been built far from city centres and their operations and consequently their competitiveness thus have always been dependent on ground transport modes connecting airports with urban areas. It means that air transport is by nature intermodal since all passengers or goods have to go from their origin point to the airport and from the airport to their destination point using ground transport modes [7]. For this reason, when dealing with airport capacity, it is necessary to consider all parts of the airport, i.e. airside (runways, taxiways and apron), landside (airport terminals) and airport ground access.

II. INNOVATIVE APPROACH TO AIRPORT LANDSIDE CAPACITY ASSESSMENT

As mentioned in the section I., the main problem of current measures that could be potentially used for mitigating the effects of the congested European air transport network is underestimation of the problems of airport terminals and airport ground access. Bearing in mind that both, terminals as well as airport ground access have direct influence on airport landside capacity we will mainly focus on addressing this issue in the following parts of this article.

Airport landside capacity assessment is very complex interdisciplinary problem that does not have a universal solution. Each airport has a specific infrastructure and is operated in a specific environment in terms of economic, geographic and demographic conditions. For this reason, it is not possible to define generic approach that could be used for assessing the landside capacity at any airport. This fact is reflected especially in the field of computer-based models of airport operations. Although these models are generally used for evaluation of the airport capacity, none of these models has attained the status of ‘international standard’ [8]. Summarizing the current status of the problems, the process of assessing the airport landside capacity is based on the set of general practices and recommendations concerning the aspects of airport operations that should be considered, and concerning the methods and tools that should be used.

The problems of airport landside capacity assessment are of industry-wide interest. Evaluation of landside capacity enables airport operators and airport designers to identify passenger and baggage flows bottlenecks, find the primary cause of the bottlenecks formation and take measures mitigating the impact of bottlenecks on the airport terminal operations. For this reason, the airport landside capacity evaluation should be an integral part of airport design and airport operations as it provides a solid base for continuous process of the airport capacity enhancement.

We have identified one principal issue in the research dealing with the problems of airport landside capacity assessment; it is the limited scope of landside capacity assessment studies. Many studies dealing with the problems of airport landside capacity are focused mainly on the processing part of the airport terminal and consider the airport terminal to be an isolated system. Although this approach provides valuable data concerning capacity, delays or processing bottlenecks, in some cases, identified capacity constraints are only the symptoms of the actual problem. In order to identify the cause of the problem, it is necessary to consider the airport terminal as an integral part of much more complex regional, national or international transportation system.

In order to solve the identified research issue, our research has been focused on investigation of the relationships between airport ground access/egress and terminal operations with a view to develop computer-based model that simulates traffic flows between passenger origin/destination and the airport.

As a result of our research and development activities a first beta version of Airport Ground Access and Egress Passenger Flow Model (AGAP) is presented in this article.

The AGAP model is a stochastic microscopic computer-based model that simulates entire airport access/egress related traffic within airport's catchment area. Its scope begins at the place of passenger's origin/destination and ends in the airport terminal. The AGAP model extends the capabilities of model simulating passenger and baggage flows in new terminal of Bratislava airport, which has been developed using PaxSim simulation tool.

Passenger movement Simulation System (PaxSim) is a set of software tools that enable simulation of passenger and baggage movements within an airport terminal and on the apron. PaxSim was developed by The Preston Group (later Preston Aviation Solutions, now Jeppesen), which is a leader in the development of advanced airspace and airport simulation, decision support and scheduling systems for the global aviation industry (The Preston Group also developed well known airside simulation tool TAAM).

PaxSim is a graphics-based computer program used for the fast-time simulation of airport landside operation. It processes information from flight schedules to determine number of arriving and departing passengers and daily distribution of traffic at the airport. PaxSim is microscopic simulation tool that allows simulating each passenger and baggage as individual objects, rather than modelling 'global' passenger flows. As PaxSim employs sophisticated algorithms of real passenger behaviour, the simulation outcomes reach a high level of conformity with real terminal operation [9].

AGAP model and PaxSim simulation model constitute a microscopic model for simulation of door-to-door passenger flows. This comprehensive simulation model enables to see the airport in the context of regional, national and international transportation network. Thanks to this approach, it is possible to analyse the interactions between traffic flows within airport's catchment area and passenger and baggage flows inside airport terminal building. This enables to identify potential capacity constraints outside the terminal building and perform comprehensive feasibility assessments of future airport ground access/egress concepts.

III. AIRPORT GROUND ACCESS AND EGRESS PASSENGER FLOW MODEL

AGAP model has been developed using MS Excel and Visual Basic programming environment. The model enables to simulate passenger flows from the place of passenger's origin (home or office) to the airport and back. Thanks to this airport ground access/egress passenger flow model, it is possible to simulate passenger flows within the airport's catchment area to and from the airport and to investigate the interactions between airport ground access/egress and airport terminal operations.

The Airport Ground Access and Egress Passenger Flow model is a stochastic microscopic computer-based model that simulates entire airport access/egress related traffic within airport's catchment area. Its scope begins at the place of passenger's origin/destination and ends in the airport terminal. The model consists of the following two modules:

- *Air passenger trips generation module:* This module is responsible for simulating the demand distribution

within the airport's catchment area. Based on the input data this module allocates passengers to particular flights, generates passenger groups and passenger distribution to the cities within airport's catchment area.

- *Passenger transport mode choice module:* This module is responsible for simulating passenger airport ground access/egress mode choice. Based on outputs from air passenger trips generation module this module selects the most favourable airport access/egress transport mode taking into account price, travel time and convenience. This module employs algorithm of passenger behaviour.

A. Air passenger trips generation module

A flight schedule is the source of primary input data for generation the air passenger trips. Before the flight schedule can be imported into the AGAP model, it has to be supplemented by additional information and all the data needs to be pre-processed to ensure they are in correct format. A completed flight schedule contains the following information on each flight: flight number, scheduled time of departure, actual time of departure, destination airports, operator, aircraft type, aircraft seat capacity, load factor, number of passengers, indication if flight is international or domestic, indication if flight is scheduled or charter, indication if flight's destination is a holiday resort, share of business passengers in flight, share of leisure passengers in flight, number of business passengers, number of leisure passengers and possible times of arrival (assuming that passenger uses services of same operator for both outbound and inbound flights)

Based on the information from the flight schedule (i.e. based on flight type, destination, aircraft capacity, load factor, and proportion of leisure and business passengers) the model allocates passengers to each particular flight. The characteristics related to passenger flows within airport's catchment area are then randomly generated and assigned to each passenger based on relevant probability distributions.

In the first step of the algorithm, the model generates the sizes of passenger groups. The air passengers often travel in groups of various sizes (e.g. families, couples, friends, business partners etc.). The group sizes are different for business and leisure passengers. Each passenger type has a probability distribution of the group size. These probability distributions are used to generate passenger groups for the flight. The algorithm generates the groups in the cycle until the number of passengers reaches the actual number of passengers in the each particular flight from the flight schedule. In the second step, the model assigns the place of trip origin/destination to each group of passengers. The region of the trip origin/destination is randomly assigned to each passenger group based on probability distribution that reflects the distribution of air transport demand within the airport's catchment area. The city of the trip origin/destination is randomly assigned to each passenger group based on the population distribution within particular region. The assignment of region and city of passenger's origin/destination is proportional. It means that if a particular region has higher air transport demand than another one, the probability that the passengers are from this region is

proportionally higher. Same analogy is used in the case of city assignment. It means that if a city within particular region has higher population than another one within the same region, the probability that the passengers are from this city is proportionally higher.

B. Passenger transport mode choice module

The algorithm of passenger transport mode choice that is used in the AGAP model is based on evaluation of the perceived costs of each transport mode. Thanks to this approach, it is possible to consider both quantitative and qualitative factors influencing the passenger mode choice. The AGAP model automatically selects for each passenger the most favourable option in terms of price, travel time and convenience.

The perceived costs of transport consist of the financial costs, time costs and transfer costs. The financial costs represent the money value needed to get from the place of origin to the airport and back including all related charges such as parking fees in case of car transport etc. The time costs represent a perceived value of in-vehicle travel time and excess travel time (i.e. waiting, walking, transfer time, etc.). The transfer costs represent a perceived value of additional physical and cognitive effort resulting from the transfer, and perceived value of risk of missing the connection.

The AGAP model evaluates perceived costs of the following airport access/egress transport modes:

- Individual car – ‘Kiss and drive’
- Individual car – ‘Park and fly’
- Taxi
- Public city transport
- National public transport + Taxi
- National public transport + Public city transport

Before AGAP model starts to calculate the perceived costs for particular airport access/egress modes, it has to calculate distances, travel times, waiting times and number of transfers for each airport access/egress option.

In the case of access/egress trips by individual cars (i.e. ‘Kiss and drive’ and ‘Park and fly’), model gathers all the required information regarding distances and travel times from the database containing comprehensive information on road network within airport’s catchment area. The time when passenger arrives at the airport before STD (Scheduled Time of Departure) of his/her aircraft is randomly generated by the model using normal probability distribution. The time when passenger leaves the airport after ATA (Actual Time of Arrival) of his/her aircraft is defined by fixed value that is estimated based on analysis of the arrival processes at particular airport.

The information related to access/egress trips by taxi are calculated and processed using same approach as in the case of individual car trips. The only difference is that in the case of taxi trip, the model randomly generates time that passenger spends by waiting for a taxi.

The information regarding national/urban public transport between particular parts of catchment area and airport are gathered from the actual public transport timetable database. The public transport timetable database contains information regarding travel times, service frequency, departure/arrival times and number of transfers for all public transport connections within the airport catchment area. The model selects the most favourable outbound and inbound connections from the database, considering the following factors:

- Passenger’s itinerary defined by departure/arrival time of his/her flight;
- Price of the connection;
- Total travel and waiting times;
- Number of transfers.

The time when passenger arrives at the airport before STD of his/her aircraft is given by the public transport itinerary of particular passenger. The time when passenger leaves the airport after ATA of his/her aircraft is given by the arriving processes at the particular airport and by time that passenger spends by waiting for the public transport connection (calculated based on the public transport itinerary).

When the model compiles a set of traffic flow related information (i.e. distances, travel times, waiting times, transfers, dwell times in terminal etc.) for each airport ground access/egress option considering a specific needs and requirements of each particular passenger, it is ready to calculate perceived costs. The value of perceived costs for all of above listed airport access/egress transport modes is calculated using the following equation (1).

$$\text{Perceived Costs} = \text{Financial Costs} + \text{Time Costs} + \text{Transfer Costs} \quad [\text{Eur}] \quad (1)$$

Assuming that airport access/egress ground transport mode with the lowest perceived costs would be the passenger’s choice the AGAP model assigns the cheapest transport option to particular passenger.

The main data output from the Airport Ground Access and Egress Passenger Flow model is an Excel spreadsheet where the information about the passengers is stored. It includes the passenger ID number, place of origin, group size, transport time etc. The most important aggregate information includes: total travel time, total distance travelled, travel costs and arrival earliness distribution of passengers.

The travel time, distance travelled and travel costs are used as performance indicators necessary for the analysis of passenger flows within the airport catchment area and for comprehensive operational, economic and environmental assessment of the airport ground access/egress solutions. The arrival earliness pattern is key information for assessing the impact of airport ground access/egress on the airport terminal operations.

The capabilities of the AGAP model have been tested on the Bratislava airport case study. Using the AGAP and PaxSim models, we have compared current design of airport ground access/egress system at Bratislava airport with innovative

concept based on dedicated minibus network serving the entire catchment area. The minibus network has been designed for collection, transportation and distribution of air passengers. Operation of minibuses within this network is based on the analogy of collection, transportation and distribution of consignments within express carriers' regional distribution network. This principle allows introducing a high-level coordination and synchronisation between air and ground transport. This airport access/egress concept is referred as *pick-up/drop-off concept* in further text.

The main aims of the Bratislava airport case study were to perform operational and environmental assessment of both airport ground access/egress concepts and to investigate their impact on passenger and baggage flows in the Bratislava airport terminal building.

IV. AGAP MODEL PRELIMINARY FORMAL VERIFICATION AND VALIDATION

The model assumptions are based on extensive passenger survey at Bratislava airport (performed during summer season in 2003, 2004 and 2007) and also on operational data collection exercise that has been performed at Bratislava airport and at Brno-Turany airport.

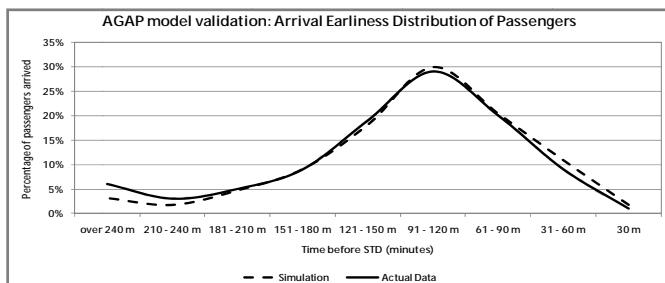


Figure 1: AGAP model validation: Arrival earliness distribution of passengers

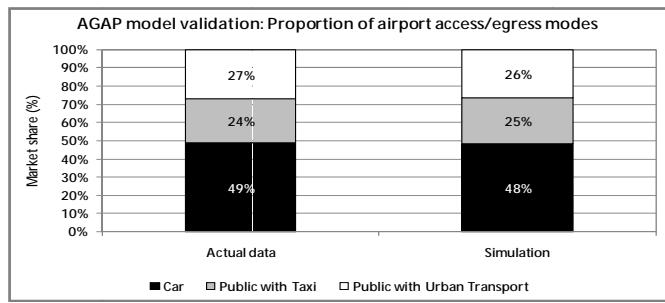


Figure 2: AGAP model validation: Proportion of airport access/egress transport modes

During the design process of the model we have created a sample of 100 passengers (randomly generated). For these 100 passengers, we have calculated all the parameters manually (e.g. group size, place of origin/destination within airport's catchment area etc.). During entire development process the functionality and accuracy of the AGAP model has been verified using this testing sample of 100 passengers. Thanks to this verification process, we have reduced the probability of creating the software bugs.

In order to validate used algorithms and verify assumptions that have been taken into account the simulation results have been compared with actual operational data. The AGAP model validation showed that the simulation results approximate the real operations. The simulation results accuracy has been verified by means of the following parameters:

- Arrival earliness distribution of passengers (see Figure 1)
- Proportion of airport ground access/egress transport modes (see Figure 2)

As can be seen from the charts the simulation results correspond to the actual operational data.

V. BRATISLAVA AIRPORT CASE STUDY

In order to perform operational and environmental assessment of the airport access/egress concept based on collection, transportation and distribution of passengers within dedicated minibus network, it was necessary to define simulation scenarios.

Traffic flows within dedicated minibus network were based on computer model of Slovak road transport network developed by Faculty of Computer Science and Management of University of Zilina. The computer model of Slovak road network infrastructure reflects various road categories and respective minibus travel speeds.

At this stage, our research did not focus on the traffic flows optimisation within dedicated minibus network. However, we assumed that by means of optimisation, it would be possible to achieve high load factors and consequently high efficiency of traffic flows.

In our study, we assumed that there is a reciprocal relationship between average minibus load factor and price per passenger-kilometre. This relationship is mathematically expressed by equation (2).

$$\text{Minibus Dist. Rate} = \frac{\text{Price of minibus}}{\text{Seat Capacity} \cdot \text{Average Load Factor}} \quad (2)$$

[EUR / km] [EUR / pax.km] [Number of Seats] [%]

The chart in Figure 3 demonstrates the price elasticity of demand for services associated with collection, transportation and distribution of air passengers.

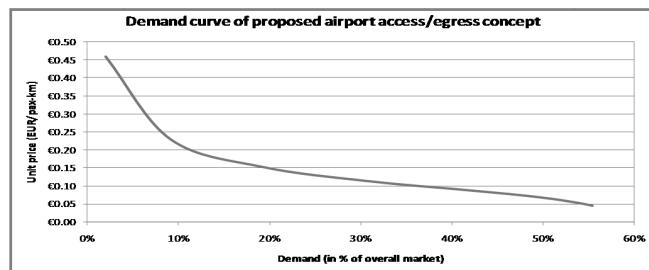


Figure 3: Pick-up/drop-off concept: Price elasticity of demand

As it has already been mentioned, average load factor is directly dependent on level of traffic flows optimisation within

dedicated minibus network. For this reason, we decided to consider three various load factor values in our simulations, in order to answer the question, what average load factor needs to be achieved through traffic flows optimisation to make the proposed pick-up/drop-off concept viable.

In all 4 scenarios, we consider the traffic flows according to flight schedule from 8th July 2008 (the busiest day in 2008). According to data that were provided by Operation Division of Bratislava airport, 49 arrivals and 45 departures of commercial passenger aircraft took place at Bratislava airport on 8th July 2008. These aircraft movements generated passenger flows of 5,497 departing and 5,900 arriving passengers, who passed through the terminal at Bratislava airport on that particular day. On 8th July, share of leisure passengers was 73% and share of business passengers was 27%.

For the purposes of operational and environmental assessment of the proposed pick-up/drop-off concept, we have defined the following 4 scenarios:

- **Baseline scenario:** This scenario considers current status of ground access/egress at Bratislava airport, without any coordination between air and ground transport.
- **Scenario40:** This scenario assumes that the proposed pick-up/drop-off concept has been introduced at Bratislava airport. This scenario also assumes that by means of traffic flows optimisation, 40% average load factor has been achieved across entire dedicated minibus network.
- **Scenario60:** This scenario assumes that the proposed pick-up/drop-off concept has been introduced at Bratislava airport. This scenario also assumes that by means of traffic flows optimisation, 60% average load factor has been achieved across entire dedicated minibus network.
- **Scenario 80:** This scenario assumes that the proposed pick-up/drop-off concept has been introduced at Bratislava airport. This scenario also assumes that by means of traffic flows optimisation, 80% average load factor has been achieved across entire dedicated minibus network.

For all 4 simulation scenarios, we assumed the following configuration of passenger and baggage processing facilities in the Bratislava airport terminal:

- **Check-in resources:** In all scenarios, we assumed 15 check-in counters opened and operated using the common check-in concept (i.e. passenger can check at any counter). In the case the baggage check-in is considered to be an integral part of the proposed pick-up/drop-off concept, 5 counters are used for self-service drop-off and 10 counters are used for classic check-in.
- **Security checks:** In all scenarios, we assumed 4 central security checkpoints to be in operation.
- **Departure passport control:** In all scenarios, we assumed 4 departure passport control counters to be in operation.

- **Arrival passport control:** In all scenarios, we assumed 4 arrival passport control counters to be in operation.
- **Baggage carousels:** In all scenarios: we assumed that 4 baggage carousels in main terminal building, and 2 baggage carousels in arrival terminal C are in operation.

In order to achieve results reflecting actual operation, we have run each scenario three times. Considering the fact that during each simulation, the AGAP model generates unique passenger sample, each scenario has been simulated and analysed using three different passenger samples. Average values of the particular outputs of these three iterations were then calculated and consequently used for further analyses.

VI. SIMULATION RESULTS

A. Market shares of the proposed pick-up/drop-off concept and other transport modes

This part is aimed at analysing the impact of proposed pick-up/drop-off concept on the overall efficiency of traffic flows within airport's catchment area.

The introduction of synchronised and coordinated airport ground access/egress is anticipated to primarily influence the proportion of particular transport modes used by air passengers. One of the principle targets of air-ground intermodality is to reduce share of individual car access/egress trips.

Proposed pick-up/drop-off concept significantly reduces market share of other airport access/egress modes. It means that pick-up/drop-off concept is able to compete with both individual and public airport access/egress transport modes. It is necessary to point out that proposed concept does not serve city of Bratislava, which is estimated to generate as much as 34.5% of the overall passenger throughput at Bratislava airport.

The simulation also showed that proposed pick-up/drop-off concept would be as attractive for leisure passengers as for business passengers:

- **Scenario40:** 29.9% of leisure passengers and 32.1% of business passengers would use the services of dedicated minibus network to travel to/from the airport.
- **Scenario60:** 47.0% of leisure passengers and 47.0% of business passengers would use the services of dedicated minibus network to travel to/from the airport.
- **Scenario80:** 53.2% of leisure passengers and 52.8% of business passengers would use the services of dedicated minibus network to travel to/from the airport.

B. Traffic flows efficiency

The simulation results also proved that proposed pick-up/drop-off concept would have a positive impact on efficiency of traffic flows within airport's catchment area. The introduction of the pick-up/drop-off concept into operation would lead to reduction of wasted times related to travelling

to/from the airport including passenger dwell times in terminal. In comparison with baseline scenario, the average wasted times related to outbound trips would be reduced by 15.6% in the case of Scenario40, by 24.9% in the case of Scenario60, and by 26.3% in the case of Scenario80. The average wasted times related to inbound trips would be reduced by 10.6% in the case of Scenario40, by 16.7% in the case of Scenario60 and by 16.3% in the case of Scenario80.

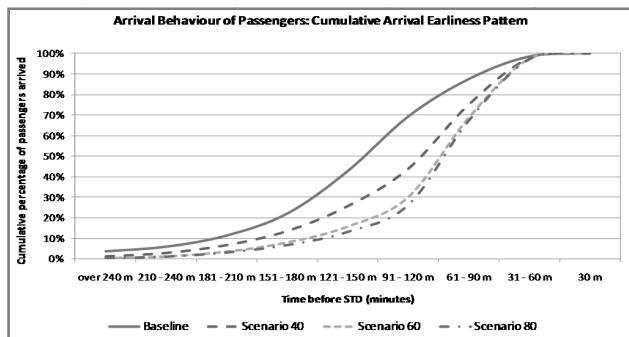


Figure 4: Cumulative arrival earliness distribution of passengers according to particular simulation scenarios.

The chart in Figure 4 depicts how the proposed pick-up/drop-off concept would contribute to the reduction of passenger dwell times in airport terminal. According to simulation outputs, in the case of Scenario40, the departing passengers would spend 23.5% less time in the airport terminal compared to baseline. In the case of Scenario60, the dwell time reduction would be 35.8%, and in the case of Scenario 80, it would be as much as 38.7%.

The fact that proposed pick-up/drop-off concept would considerably contribute to increased efficiency of door-to-door transportation is reflected in a significant reduction of average number of transfers per access/egress trip. The average number of transfers per access/egress trips would be reduced by 44.1% compared to baseline in the case of Scenario40, by 69.0% in the case of Scenario60, and by 72.0% in the case of Scenario80.

On the other hand, faster and more convenient transport mode would be more expensive in terms of direct financial costs. However, the advantages in terms of convenience and time savings surpass higher travel costs. It means that proposed pick-up/drop-off concept is still cheaper in terms of perceived costs.

C. Environmental assessment

The simulation results that has been analysed in previous chapters proved that the proposed pick-up/drop-off concept is able to compete with private cars in terms of travelling speed and convenience, and thus contribute to the reduction of share of individual car access/egress trips. Consequently, if the proposed pick-up/drop-off concept reaches certain market share it could also contribute to the reduction of air pollution related to ground traffic generated by the airport.

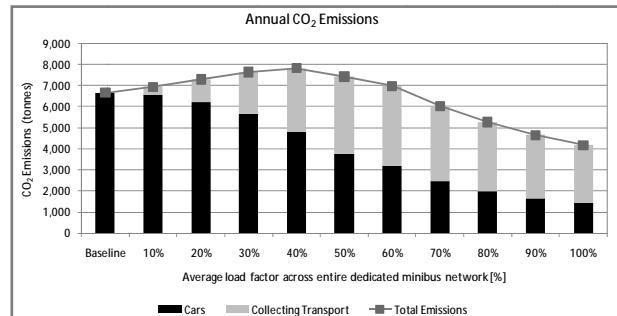


Figure 5: Annual CO₂ emissions caused by individual car access trips and by dedicated minibus network operation

However, the fast-time simulation of traffic flows within airport's catchment area showed that significant reduction in CO₂ and SO₂ emissions (for CO₂ see Figure 5) would be achieved only if average load factor reaches 70%. In terms of NO_x emissions, the introduction of the proposed airport access/egress mode would lead to their increase regardless the minibus traffic flows efficiency. This results from the fact that diesel minibuses produce significantly more NO_x emissions compared to commonly used cars. On the other hand, we expect that even lower intensive utilisation of minibus fleet would contribute to reduction of local air pollution related to cold starts of private cars.

D. Airport terminal operations

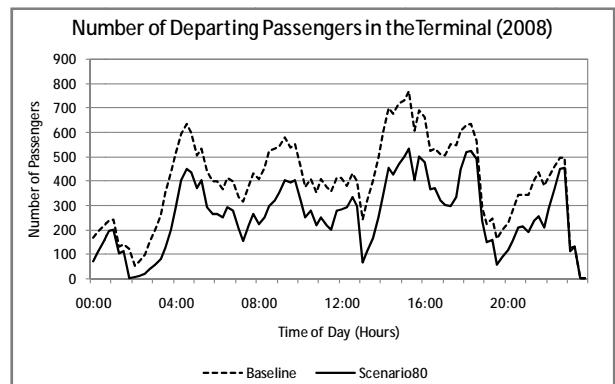


Figure 6: Number of departing passengers in airport terminal building during the day

Fewer passengers in terminal pose fewer requirements on the size of the airport terminal building and thus increasing investment efficiency. According to the simulation outputs, introduction of synchronisation between airport ground access and airport traffic would lead to significant reduction of number of departing passengers in the terminal. In the case of Scenario40, the average number of departing passengers in the terminal would be 21.0% lower compared to baseline. In the case of Scenario60, the average number of departing passengers would be reduced by 33.0%, and in the case of Scenario80, the average number of departing passengers would be reduced by as much as 35.7% (for baseline and Scenario 80 see Figure 6).

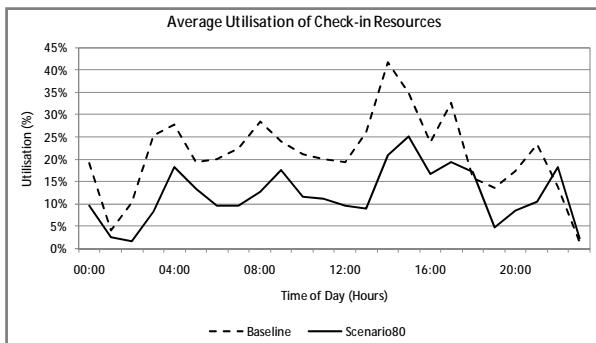


Figure 7: Average utilization of check-in resources

The average utilisation of terminal processing resources would only be affected if the baggage check-in is an integral part of the proposed pick-up/drop-off concept. Moreover, this applies only to utilisation of check-in resources. The impact of the pick-up/drop-off concept on utilisation of other processing facilities is insignificant. According to simulation results, if the baggage check-in is an integral part of the proposed pick-up/drop-off concept, it would be possible to handle same number of passengers using 25.4% less check-in resources in the case of Scenario40, 37.8% less check-in resources in the case of Scenario60, and 43.2% less check-in resources in the case of Scenario80 (for baseline and Scenario 80 see Figure 7).

The impact of the proposed pick-up/drop-off concept on other terminal processing facilities (i.e. security checks, passport control counters, etc.) is insignificant.

VII. CONCLUSIONS

This paper describes and demonstrates a new method for evaluating the capacity of airport terminals as well as for operational and environmental assessment of airport ground access/egress system. This new method is based on fast-time simulation of door-to-door passenger flows and thus enables to see the airport terminal as an integral part of regional, national or international transportation network. Thanks to this fact, it is possible to analyse the interactions between airport ground access/egress and passenger and baggage flows inside airport terminal building. The new method reveals an innovative approach to performing comprehensive operational and environmental assessments of future airport ground access/egress concepts.

Using this new approach, we have performed an operational and environmental assessment of innovative airport access/egress concept based on the intermodality principles that are widely used within integrated intermodal networks of parcel companies. Thanks to microscopic simulation of door-to-door passenger flows we were able to conduct initial feasibility assessment of the proposed pick-up/drop-off concept and identify its potential benefits.

VIII. FUTURE WORK

At this stage of research and development it is not possible to use developed simulation models as decision making support

tools in real operations. It is necessary to perform a more comprehensive validation of outputs.

Within further research, we will also focus on the following issues:

- Development of more sophisticated algorithms of passenger transport mode choice (e.g. current model assigns each passenger with the cheapest transport option, which does not fully reflect the actual passenger preferences);
- Integration of algorithms reflecting the probability of delay in both, air and ground transport.

ACKNOWLEDGMENT

We would like to express our gratitude to Jeppesen (former Preston Aviation Solutions) for providing us with academic license of PaxSim simulation tool that enabled us to develop computer simulation model of new Bratislava airport terminal operations.

REFERENCES

- [1] EUROCONTROL STATFOR, "EUROCONTROL Long-term forecast: IFR flight movements 2008 – 2030"; EUROCONTROL, 2008
- [2] EUROCONTROL, "Challenges to growth 2008 summary report"; EUROCONTROL", November 2008
- [3] B. Tether, S. Metcalfe, "Horndal at Heathrow? Co-operation, learning and innovation: Investigating the processes of runway capacity creation at Europe's most congested airports", The University of Manchester, UMIST, June 2001
- [4] A. Evered, "The impact of new security measures on airport retailing in the UK", Airport International website (www.airport-int.com), 2007
- [5] A. Kazda, R. E. Caves, "Airport design and operations", Second Edition, Elsevier, 2007
- [6] EUROCONTROL, "Challenges to growth 2004 report", EUROCONTROL, December 2008
- [7] M3 Systems, ANA, ENAC, LEEA, "CARE II: The airport of the future: Central link of intermodal transport?", WP1: Review of current intermodality situation, EUROCONTROL Experimental Centre, October 2004
- [8] R. Neufville, A. R. Odoni, "Airport systems: Planning, design and management", McGraw-Hill, 2003
- [9] Preston Aviation Solutions, "PaxSim solutions user manual", Preston Aviation Solutions Pty Ltd., 2006
- [10] M. Stefanik, "Problems of airport capacity assessment", Doctoral thesis; University of Žilina, 2009
- [11] M. Bugaj, "Fault tree analysis", Studies of Faculty of Operations and Economics of Transport and Communications of University of Žilina, page 23 – 27, 2004
- [12] L. Beno, M. Bugaj, A. Novak, "Application of RCM principles in the air operations"; Komunikácie – vedecké listy Žilinskej univerzity; ročník 7, číslo 2, 2005, page 20 – 24
- [13] B. Badanik, Airlines' point of view as a new approach to measuring quality of service at airports, ICAS 2008 (26th Congress of the International Council of The Aeronautical Sciences), Anchorage, Alaska, USA, 14 – 19 September 2008
- [14] N. Ashford, H. P. M. Stanton, C. A. Moore, "Airport operations", Second Edition, McGraw-Hill, 1997
- [15] A. Cokasova, "Analysis of passenger viewpoints and of the practical shift in air rail intermodal transport", Doctoral thesis, University of Žilina, 2006
- [16] IATA, "Airport development reference manual", 9th Edition, IATA, 2004

Operational Evaluation of an Airport Centered Flow Management

Eike Rehwald & Peter Hecker

University of Braunschweig
Institute of Flight Guidance
Braunschweig, Germany
e.rehwald@tu-bs.de & p.hecker@tu-bs.de

Rainer Kaufhold

DFS GmbH, German Air Navigation Services
Tower Management Services - SESAR
Langen, Germany
Rainer.Kaufhold@dfs.de

Abstract – As hub airports become larger and larger, it is vital that available runway capacity is used optimally to prevent them turning into air traffic bottlenecks. This paper presents the Cooperative Local Resource Planner (CLOU), which has been developed as a prototype to assist in “airport-centered flow management”. An overview of the first steps to be taken to guarantee a smooth operational implementation is also given. Different runway-use strategies will be discussed, using the German Frankfurt Airport as an example. Furthermore, the display of the planning results of CLOU and the integration into the air traffic controller work area are addressed. Finally, embedding of CLOU into existing system environment is presented.

Keywords: Air Traffic Flow Management, Network and Strategic Traffic Flow Optimization, CLOU

I. INTRODUCTION

Nowadays the European central hub airports often operate at their capacity limits (compare [1] and [2]). More and more, they are becoming the bottlenecks of the air transport network. Even today, the smallest incident (which might either be a reduction of available capacity or a shift in demand) at a hub airport can cause huge delays and adversely impact operating efficiency. These impacts are not limited to the operations of a single airport, but can negatively affect the complete European airspace in terms of a “reactionary delay” (compare [3]).

Expanding a hub airport results in complex runway systems, with complex interdependencies between the runways. These interdependencies result either from mixed-mode operations or from interactions with the adjacent airspace. Despite such expansions, it can be assumed that capacity bottlenecks will remain an issue, at least at traffic peaks.

To use the capacity of the runways optimally, a Flow Management System as a Cooperative Local Resource Planner (CLOU) has been developed at the German Aeronautical Research Program sponsored by the Federal Ministry of Economics and Technology of the German Government. It provides suggestions for the chronology of runway-use strategies based on demand and capacity prognoses. After a detailed technical examination of the system with live data, the supervisors of Tower and Approach Control at the German hub Frankfurt Airport (EDDF) will now perform an operational validation of the prototype.

CLOU is a database-based airport-centered flow management tool, which extracts flight information of Stanly_CDM and INFO+ (via Capacity Manager (CAPMAN, calculates and forecasts the available airport capacity) and only at Frankfurt Airport). With a planning horizon from 30 min up to six hours CLOU fills the gap between tactical systems like Arrival Manager (AMAN)/Departure Manager (DMAN) and pre-tactical network planning like Central Flow Management Unit (CFMU, operational unit of EUROCONTROL). Based on demand and capacity considering constraints and optimization parameter CLOU generates a prediction of expected runway in use, an optimal operation procedure, runway workload, and parameters every five minutes. These parameters are flow, punctuality, adherence to schedule, delay, and queue. The planning results of CLOU are shared with other prediction tools like CAPMAN (compare [9]). The algorithms that are used to optimize the runway-oriented flow management have already been presented before (compare [5] and [6]) and ain't be part of this paper.

In addition to the underlying idea behind the flow management of complex runway systems concept, this paper also presents the first results of the operational validation. These results emphasize the challenges presented by the integration of this concept into operational procedures.

II. PRESENT SITUATION

At airports with runway systems, it is possible to handle flights over different runways. However, as a rule, all departures with the same destination direction leave from the same runway, since non-systematic runway assignment can quickly result in confusing situations in the airspace. Unfortunately, this may mean that one runway is overloaded, while there is unused spare capacity on another. A better balance can be achieved by shifting departures, grouped by Standard Instrument Departure (SID), among the different runways.

Fig. 1 shows an example of such an operational procedure for Frankfurt Airport. In addition to the dependent parallel runway system 25/07 for arrivals and departures, Frankfurt Airport also features runway 18, or ‘runway west’, for departures only.

At Frankfurt Airport, flights leaving to the north and northwest usually take off from the parallel runway system. Departures to the south, west and east are assigned to runway

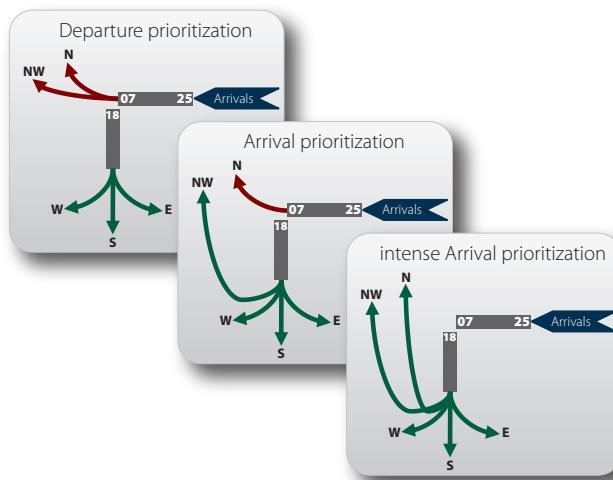


Fig.1. Example of operation procedures at Frankfurt Airport

west. Arrivals are handled exclusively by the parallel runway system.

To ensure adequate arrival capacity during an arrival peak on runway 25/07, it is possible to move either northwest or north and northwest departures to runway 18.

The supervisors on duty take the decision to relocate the departure flow from one runway to another based on a personal assessment of the situation. This does not pose a problem as long as the alternative runway has enough capacity to handle the additional departure flow without causing delays. However, as a rule, this decision is not so trivial. An assessment must be made as to whether any resulting delay from relocating the departure flow is indeed less than the delay from using the standard runway.

The decision-making process is further complicated by the necessary negotiations between Tower supervisors and Approach supervisors. Naturally, Tower supervisors focus on departures, whereas Approach supervisors prioritize arrivals. The fact that Tower and Approach Control belong to different DFS business units makes a holistic point of view rather difficult.

This is also reflected in the systems currently available to support supervisors in their decision-making. The 'arrival manager' controls inbound traffic without taking departures into account, while, on the departure side, a flight data processing system is used which does not take inbound traffic into account. The long-term planning of CFMU does not allow a holistic view of the traffic processing at airports either.

The situation is becoming more and more complex, as international hub airports add new runways, which result in ever increasing interdependencies among the runways. The optimal utilization of the available capacity over the daily peaks in inbound and outbound traffic is just the start. Further factors that must be optimized include the impact of weather and noise abatement procedures on runway operations.

III. PURPOSE: FLOW MANAGEMENT

The purpose of CLOU is to optimize the traffic flow of an airport's runway system. CLOU supports coordinated decision-making between Tower and Approach as regards the prioritization of both arrival and departure traffic. It provides suggestions for optimal runway-use strategy and the point in time to change strategy.

For example, during an arrival peak at Frankfurt Airport, CLOU might suggest shifting the departure flow from the normally used parallel runway system to runway west to minimize the overall delay.

In addition, CLOU proposes a prioritization of the remaining arrivals and departures on the parallel runway system. The planning suggestions generated by CLOU are based on a dynamic capacity and demand prognosis, taking into account interactions between in- and outbound air traffic. The surface and turnaround flight phases are reproduced by simple logic and flight information.

Supervisors can manually enter previous experiences into the system or visualize the flow behavior following a proposed change in strategy. Hence, CLOU provides a basis for discussion for a more collaborative decision-making process among supervisors.

IV. DISPLAY/VISUALIZATION OF PLANNING RESULTS

CLOU displays the planning results which allow supervisors to see basic planning suggestions for a planning horizon of three or more hours. Additional tab sheets provide access to more detailed information on individual flights or such calculated parameters as flow, punctuality and delay.

The results are assigned from left (actual point of time) to right (increasing planning horizon), divided into ten-minute intervals (see Fig.2). The bottom line "time" shows the UTC time in half-hour increments.

The first row contains the prognosis for the expected runway-in-use. This information is based on the weather forecast and may be manually modified by the supervisor.

The suggestion regarding which runway-use strategy to apply is shown in the second row "DEP 25/07". Every runway-use strategy is assigned a color and a designator.

The third row illustrates the overall capacity of the runway system, presenting the basis for the optimization. This information is supplied by the airport system CAPMAN (operated by Fraport, the operator of Frankfurt Airport). If a supervisor judges the available capacity to be greatly different, is it possible to replace the values manually as well. This option exists both for overall capacity and for partial capacity (individual runways or individual capacity of arrivals or departures).

The visualization of arrival or departure prioritization follows in the rows below ("Rwy 25/07" and "Rwy 18"). The number of flights per ten-minute interval is color-coded – using different colors for arrivals and departures – as well as bar graph-coded. Departures act like stalactite, in contrast arrivals behave like stalagmite. Furthermore, the caused delay by the ten-

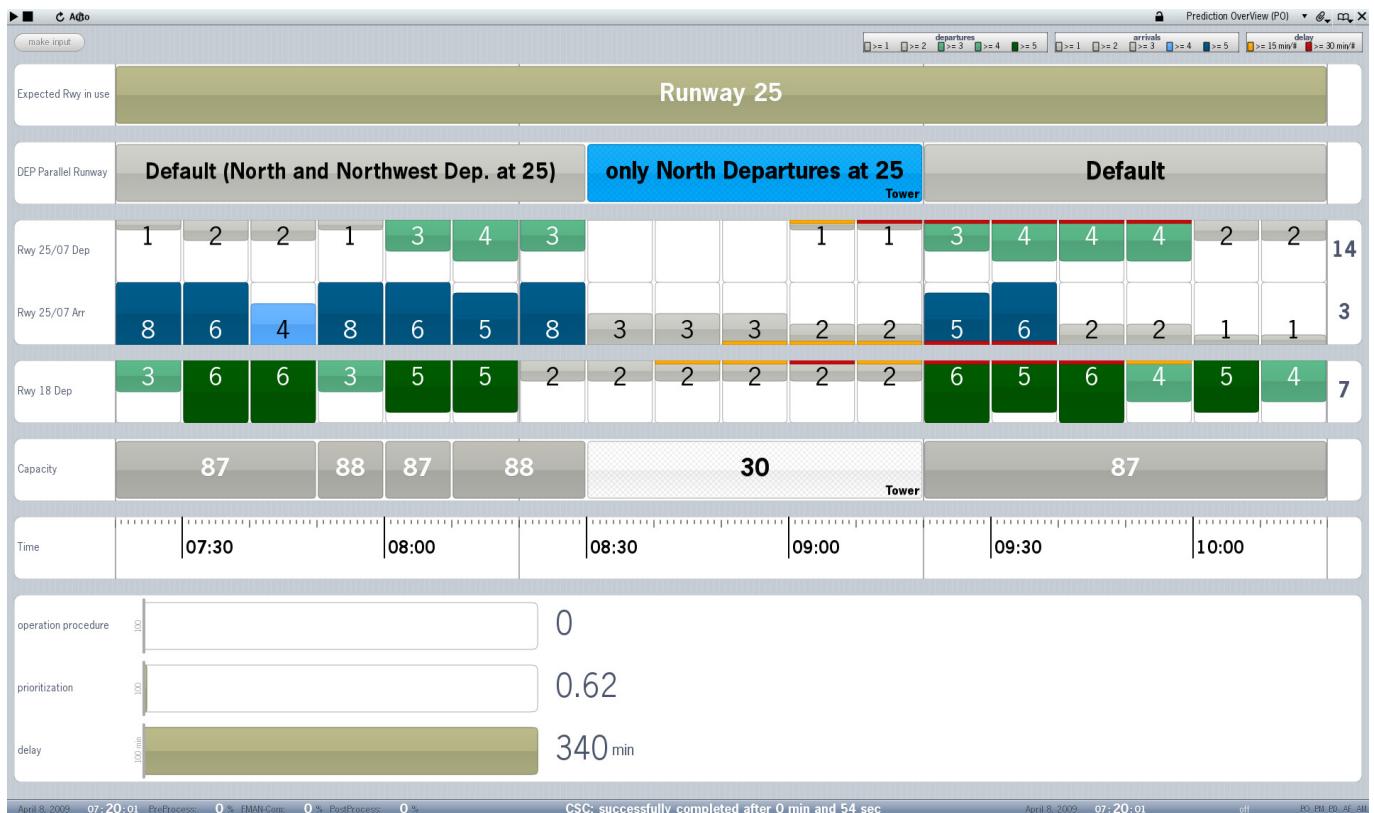


Fig.2. Visualization of planning results for supervisors

minute interval is shown as two steps over 15 minutes and over 30 minutes (limits are adjustable). As additional information, the average delay per flight during the shown intervals and according to the runway are displayed on the right-hand side.

The bar graph at the bottom pictures the difference between delay of the initial and flow optimized result. It is presented in minutes and contains the amount of all flights.

The visualization means that supervisors can create a mental picture of the optimized traffic situation, which in turn, provides a basis for discussion. It should be borne in mind that the suggestions are not binding; they merely provide decision-making support for supervisors. The system should not replace the supervisor's decision, as situations may arise where the system does not have all necessary information to create an optimized solution.

V. OPERATIONAL EVALUATION

In September 2008 and from May through July 2009, first tests within the operational environment were carried out (compare [7] and [8]). This will be followed by a field test scheduled for the second quarter of 2010. One of the aims of the field tests was to allow supervisors of Tower and Approach to evaluate the usability of CLOU in an operational environment. Furthermore, it has determined any additional requirements that are still lacking from the supervisors' point of view. At this point, this paper will present first results and problems of the operational tests. The design of the supervisor's working position during the field test will be covered, including any changes needed. The steps necessary to increase supervisor acceptance of the system will be addressed.

A. Integration into the working position

The algorithms behind CLOU have been tested with live data. For this purpose, an internal network with live data access was created. The next step is to test the usability of the prototype in an operational setting. The question arises how best to integrate the new planning information into the supervisor's working position.

CLOU is still a prototype which means that it is not possible to integrate the planning information into a live operating system. In addition, severely limited space means that it is not possible to set up an additional monitor at the supervisor's workstation. There is generally no free space available in the Tower and the supervisor workspace in Approach Control is already filled with various monitors so that there is no room for a new display there either. Hence, an additional screen is neither reasonable, nor realizable.

Therefore, a different approach was taken for the first tests. The CLOU computer itself remains in the research laboratory and the planning data was exported via intranet to a computer that is not connected to any operating systems. This test arrangement presents the only means of performing an operational evaluation within the means available. One of the disadvantages of this solution is the fact that the chosen screen is also used to retrieve other information, so that CLOU can not be displayed all the time.

After analyzing the information from field tests, the best method to integrate CLOU into the existing working position will have to be determined in cooperation with engineering and operational staff.

B. Change Management

During the development phase of the CLOU prototype, the operational staff contributed by describing the various operating procedures and the interdependencies between in- and outbound traffic. Usually, real-time simulations using operational personnel are conducted to assess the user benefit of new prototypes. However, this approach could not be taken because CLOU is a system supporting pre-tactical work.

In this case, an alternative procedure was adopted and in-depth discussions about the optimization concept were held with Tower and Approach supervisors. These discussions not only helped clarify the need for a support system, they also highlighted the supervisors' reservations about such a system. This underscored once again the importance of a carefully planned implementation of the new system.

The three main reservations of the supervisors and the suggested solutions are described below.

1) Trust in planning systems

During the first tests and the discussions with the supervisors, it became clear that air traffic controllers harbored general doubts about planning systems such as CLOU. These doubts result from experiences with the introduction of various planning systems in the past and lack of knowledge of the new system.

During previous implementations of different planning systems, the role of change management had been underestimated. Staff were often instructed to strictly adhere to the decisions and suggestions produced by such systems, although they had no background information about the underlying processes. They did not know what basis the system used to produce its decisions. Hence, the staff could not develop the necessary trust in the system's reliability. The fact that decision-making was taken away from the supervisors and given to a system with an unclear mode of operation resulted in the complete rejection of such systems.

CLOU will run in parallel with the other systems, without the need for extra inputs from the air traffic controller. The system updates itself every five minutes with new initial data. The results are presented as a suggestion to the air traffic controller. The air traffic controller may then use this information to evaluate his decision-making process. The air traffic controller may possess additional information not included in CLOU. If the air traffic controller makes a decision that is not in line with the system's suggestions, CLOU will automatically update the initial setting of the flight plan data. The air traffic controller can also directly enter the information or decision into CLOU.

2) Transparency of optimization

Nowadays, air traffic controllers concern themselves mainly with their own sector. A consideration of the long-term traffic situation or the situation in neighboring sectors does not yet take place. With CLOU, the controller's view of the air traffic situation is enlarged. The optimization process that is working in the background does not only consider Approach or Tower prioritizations, but also the best compromise for all airport sectors. It is important that the air traffic controllers do not only focus on the technical output of some optimization algorithm, but in fact change their way of thinking overall.

When air traffic controllers understand the importance of the optimum solution for the overall situation, they will be

more likely to accept a potential temporary worsening of the situation in their own sector if called for.

Besides developing trust in the system's ability to find an optimal overall solution, it is also essential to create outputs that present the situation and suggestions explicitly. Air traffic controllers will have to get used to the display of the results. If they feel comfortable and familiar with the display, they will extract the necessary information from the display quickly and without any hesitation or doubt.

By introducing punctuality as an optimization criterion, air traffic controllers will have to develop the capability to evaluate the present traffic situation accordingly. For air traffic controllers, it is rather difficult to categorize a flight as punctual or unpunctual without the help of systems like CLOU. Therefore, such flow management systems are necessary when introducing new means to air traffic optimization.

3) Adequate data quality

Adequate data quality is the linchpin of the entire flow management. An optimization is only as good as the initial data quality. But of course it is also possible that the system is missing some boundary conditions, such as information on the reduced flow in preceding sectors. In order to avoid an optimization based on false data, CLOU must have the capability of manual input, allowing supervisors to modify parameters manually.

Therefore, the CLOU interface contains a tab sheet to allow the modification of parameters with minimal effort (see Fig.3). The interaction tab sheet appears in the same design as the tab sheet display.

The following parameters may be changed: runway-in-use, operations procedure, overall capacity, partial capacity of runways as well as numbers of arrivals and departures per runway.

Inputs in the tab sheet will be recognized automatically by CLOU and trigger the refreshing of the optimization to guarantee results that are always up-to-date.

The ability to change parameters manually introduces a new requirement. Air traffic controllers are not used to handling direct capacity values, although these are needed for CLOU.

VI. EMBEDDING INTO EXISTING SYSTEMS

The general concept of "balancing of demand and capacity" is not new at all. Based on demand and available capacity, target times are generated for every single flight. Nowadays,

Expected Day & Runway	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	+	
DEP Parallel Runway	d	d	d	d	d	d	N	only North Departures at 07 prototype					d	d	d	d	+	
Cap Dep 25/07	24	24	24	25	24	25	25	8	8	8	8	8	24	24	24	24	24	+
Cap Av 25/07	52	52	52	53	52	53	53	18	18	18	18	18	52	52	52	52	52	+
Cap Dep 18	37	37	37	37	37	37	37	13	13	13	13	13	37	37	37	37	37	+
Capacity	87	87	88	87	88	88	30	prototype					87	87	87	87	87	+
Time	07:30		08:00		08:30		09:00		09:30		10:00							

Fig.3. Interaction tab sheet

this happens with the help of the CFMU – a pre-tactical system – as well as with the help of tactical systems, such as the arrival (AMAN) and departure managers (DMAN). The main difference between pre-tactical and tactical planning is the increasing accuracy of the boundary conditions and hence improved planning quality.

CLOU closes a gap both between pre-tactical and tactical systems, as well as in respect of coordination of the interaction between in- and outbound traffic at an aerodrome.

The different levels of “balancing of demand and capacity” have different goals, as depicted in Fig.4.

A. CFMU

The CFMU aims to avoid overload within sectors. The focus is on the approach sectors in this case. CFMU is comparable with an open-loop control. Flights are assigned with slots but no update is carried out during traffic handling.

Furthermore, the CFMU is a network planning system with no special focus on airports.

B. CLOU

With CLOU, a changeover to “closed-loop” control will be introduced to air traffic control. CLOU distributes the demand among the available runways and assigns priorities between in- and outbound traffic. By keeping the system updated with the newest traffic information, CLOU ensures a permanent ongoing balancing of demand and capacity.

CLOU is an airport-oriented system that also considers network issues.

C. AMAN/DMAN

Arrival and departure managers concentrate on minimizing separation, as well as on the coordination between air traffic controllers. This tactical system is arrival-oriented only and provides current times.

VII. PRELIMINARY RESULT

During field tests so far the prognosis of expected-runway in use was very well. With manually inputs it was possible to define an explicit time to change operation direction.

The capacity forecasts was reliable as well. But exceptional cases have needed manually input, for example, borderline tailwinds at runway west combined with pilot decisions.

Based on these results and an adequate data quality, air traffic controller review the suggestions of operation procedure and runway workload. As basis of decisionmaking appears three characteristics: number of shifted flights, value of delay improvement, and forecast stability.

1) Number of shifted flights

One reason to refuse the suggestion is only a small number of shifted flights between runways. The fairly low delay improvement doesn't justify the accelerated coordination effort.

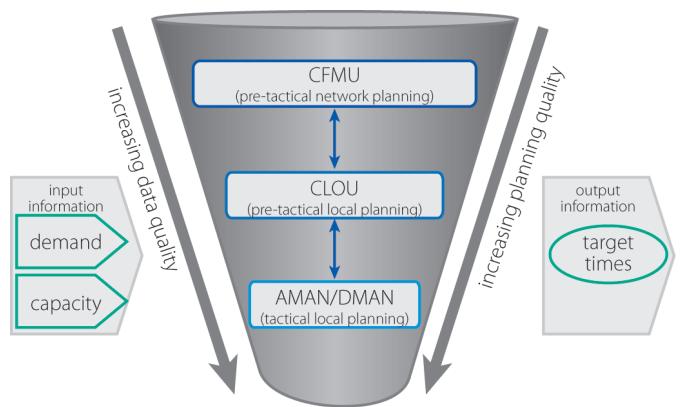


Fig.4. Levels of “balancing of demand and capacity”

2) Value of delay improvement

As main basis of decisionmaking emerges the value of delay improvement over the planning horizon. This value is visualized as bar graph (compare Fig.2). In case of a delay improvement of over one hundred minutes between the initial first-come first-serve and the flow-optimised result based on an optimal operation procedure, the air traffic controller follow the suggestions of CLOU, usually. In this instance an improvement of the traffic situation was noticeable.

3) Forecast stability

It turns out, that a forecast stability has to be guaranteed. For example an operation procedure switch takes up to half an hour. From this point an operation procedure forecast has to be stable the next hour. Therefore, an optimal response due to traffic changes is only aggravated possible.

A. Restrictions of field tests

Due to the fact that CLOU is a prototype, all supervisors and air traffic controller were asked to have a look at CLOU and review the suggestions with their own expertise. From it, they are free to follow the suggestion and to prove it. But this is volunteer in doing so. The air traffic controller accounts for his decision.

B. Air traffic controller résumé

After these two first field tests a mainly positive response of air traffic controller is noticed. The estimated benefit of CLOU with the actual airport topology is elusive from the air traffic controllers point of view. But with the upcoming four-runway-layout according to the much higher runway complexity, air traffic controllers expect a noticeable benefit with CLOU.

VIII. CONCLUSION

The development of the algorithm in the CLOU prototype is nearly finished. Research with live data has proved that CLOU has the potential to reduce delay and at the same time improve punctuality. To validate the system, field tests are indispensable. These experimental tests have shown in which way the provided information represent helpful support for the air traffic controllers concerns.

With the help of flow management, capacity bottlenecks at hub airports can be detected in a timely fashion, allowing to take corrective action much earlier than at present. This means that not only the airport that uses CLOU profits from the system, the situation in the surrounding airspace is relieved as well.

The CLOU interface informs the controller about the future air traffic situation and about a possible solution for the runway-use strategy. Based on this information, Tower and Approach could agree on further procedures and record them per input into CLOU.

The field tests offer the possibility to get a first validation by air traffic controllers during operations. Furthermore, air traffic controllers are able to voice constructive criticism and make further suggestions concerning the functions of CLOU and its human-machine interface. These points will be considered in the further development of CLOU.

Furthermore, field tests should indicate whether the optimization results should be given to the supervisors only, or if they could also be of help to the air traffic controller. Normally, a supervisor does not deal with individual flights. This is part of the controller's duties. Therefore, it seems reasonable to provide the air traffic controller with the results of CLOU.

IX. OUTLOOK

On the basis of the runway-related demand and capacity forecasts, further applications of CLOU will be developed both within the framework of the German national research program "Innovative Airport (iPort)" and the SESAR initiative. These include:

A. Optimization regarding punctuality

By using a modified objective function, traffic handling can be optimized to also take account of punctuality instead of just aiming at minimizing delays as is the case today.

B. Use by airlines and airports

Particularly in the case of major problems in traffic handling (reduced capacity or shift in demand), airlines and airports will be better informed about the effects of such disturbances with regard to delay and punctuality. They will thus be in a position to plan their processes (aircraft turnarounds, parking positions, etc.) with longer lead times in a proactive instead of a reactive manner.

C. Implications for the CFMU

By considering arrivals and departures as integrated processes over a longer lead time, the CFMU will be more precisely informed about time changes. The CFMU can thus adapt CFMU slots as necessary and ensure better use of airspace in analogy to the early take-off time used in A-CDM.

D. Prioritization by means of AMAN/DMAN

Traffic handling can be further optimized by combining the systems CLOU and the sequence-oriented planning of the arrival and departure managers.

E. Target time management

Thanks to the long lead times, optimum management of individual flights can be initiated at an early stage taking airline preferences into account. Reliable planning of departure and arrival times is essential for the future 4-D trajectory management since start and end of a trajectory are defined by these times.

REFERENCES

- [1] "Air Transport in Germany – Mobility Report 2007", "Luftverkehr in Deutschland – Mobilitätsbericht 2007", DFS Deutsche Flugsicherung GmbH, 2008
- [2] Eurocontrol, Medium-Term Forecast – Flight Movements 2008-2014, Vol. 1, Eurocontrol Statfor, February 2008
- [3] Eurocontrol Trends in Air Traffic, A Matter of Time: Air Traffic Delay in Europe, Vol. 2, Eurocontrol Statfor, September 2007
- [4] R. Bruder, C. König, T. Hofmann, and H. Rabenstein, "Zwischenbericht – Arbeitswissenschaftliche Leistung im Rahmen des Verbundvorhabens Wettbewerbsfähiger Flughafen", unpublished
- [5] R. Kaufhold, K. Nachtigall, S. Marx, and C. Müller-Berthel, "A pre-tactical Generalised Air Traffic Flow Management Problem", Paper, ATM Seminar 2007
- [6] A. Pick, "Validation results of Airport Total Operations Planner Prototype CLOU", Paper, ATM Seminar 2007
- [7] R. Kaufhold, "CLOU Bewertung – Test im operationellen Betrieb", unpublished
- [8] R. Kaufhold, Ch. Hüber, "Auswertung CAPMAN/CLOU Feldversuch", unpublished
- [9] Fraport AG, "Optimum Capacity Utilization CAPMAN – Capacity Manager"

Door-to-Gate Air Passenger Flow Model

Martin Matas
 Air transport department
 University of Zilina
 Zilina, Slovakia
 martin.matas@fpedas.uniza.sk

Milan Stefanik, PhD,
 milan.stefanik@fpedas.uniza.sk
 and Sandra Kollova, PhD
 sandra.kollova@fpedas.uniza.sk
 Air transport department
 University of Zilina
 Zilina, Slovakia

Abstract—Ever growing traffic in air transport with associated capacity constraints brings problems to air passenger flows at airports. In efforts for improvement new original future airport concepts are thought out. For the purpose of evaluation of future airport concepts the passenger flow model is developed. The model consists of two sub-models: Airport Ground Access Passenger Flow Model (AGAP) and Airport Terminal Passenger Flow Model (PaxMod). AGAP is based on random generation of passenger flows from the catchment area to Airport Bratislava using statistical data. PaxMod is based on linked cumulative diagrams representing airport queuing systems and simulates passenger flows through the airport terminal facilities. Both models are interconnected and are used to evaluate Airside-Landside Separation concept (ASLS) by simulating two scenarios. First scenario is baseline scenario where classic air passenger transport is simulated. Second scenario simulates passenger flows in Airside-Landside Separated airports and the result of simulation is compared to the baseline scenario. Simulations showed that for most passengers the door-to-gate transit time in ASLS scenario is higher than in classic scenario.

Keywords-Passenger Flow Model, Airport Terminal, Airport Access, Queuing, Cumulative Diagrams, Travel Time, Airport Catchment Area, Air Passenger

I. INTRODUCTION

The Door-to-Gate Air Passenger Flow Model is developed for the design and evaluation of original future airport concept of Airside-Landside Separation which idea was described in [1] and [2]. It is able to simulate the passenger flows from their homes through the airport catchment area and the terminal to the airport gates. The passenger flows at airports in this model consist of processes (check-in, security control, boarding) and movements among the processes.

The air passenger processes can be modelled by analytical queuing models (stochastic or deterministic) or by simulation models. In [3] an extensive survey on passenger behaviour at Manchester Airport was made for the purpose of developing an analytical model of passenger time spent at the airport. The model is based on a network of linked analytical queuing models where the nodes represent the processing centres, and the links represent the proportion of total passenger flow. Alternatively to stochastic approach [4] proposed deterministic queuing models which could be graphically analysed by cumulative diagrams as in [5]. This approach is used in [6] to

model passenger arrivals to the departure lounge and their departure from the lounge to the aircraft. The proposed deterministic function describing cumulative passenger arrivals was a quadratic function. Simple landside aggregate model presented in [7] is an analytical aggregate model for estimating capacity and delays at airport terminals. The facilities in the terminal are divided into three classes: processing facilities, holding and flow facilities. In processing facilities passenger dwell time is calculated using deterministic equivalent queuing model. Analytical models can be used to study impacts of certain parameters on the system. On the other hand to keep their underlying equations tractable they are often based on strong assumptions which tend to be unrealistic. If the system becomes too complex for analytical modelling the simulation models might become preferable. The simulation model of the complete passenger flow from the check-in to boarding and from de-boarding to baggage claim was modelled in [8]. This model and other models of airport terminals presented in [9] and [10] were simulated using ARENA simulation software. Although many authors develop their own simulation tools [11], there exist specialized tools for passenger and baggage flows at airports such as PaxSim.

In the context of our research the air passenger movement at airport terminals is regarded as passenger walking. Walking behaviour can be analysed on a different level of detail (microscopic, mesoscopic, macroscopic) and using different modelling techniques or theories. In our literature survey models are classified according to modelling approach to the following classes: Microsimulation models, Cellular Automata models, Queuing theory based models, Gas-kinetics based models and Continuum physics based models. This classification has been adopted from [12].

For the purposes of our modelling we are interested in passenger flow as a whole rather than in individual passengers. However we still want to distinguish different types of passenger groups. In particular we are interested in the classification of passengers to business and leisure and their corresponding flights such as long-haul vs. short-haul, scheduled vs. charter, domestic vs. international and so one. The analytical stochastic queuing models have difficulties in capturing the quickly changing passenger arrival rates at airport check-in desks or airport gates. The discrete-event micro-simulations tend to be too complex and require a lot of input data. Therefore we decided to use a simulation approach based

This research is conducted thanks to the support from Eurocontrol Experimental Centre (Bretigny sur Orge, France) in cooperation with University of Zilina (Zilina, Slovakia).

on linked deterministic queuing models for modelling of passenger flows at airport terminals. The airport ground access flows are modelled by random numbers generation based on probabilistic distributions of passengers within the airport catchment area and by assigning them the transport mode with the lowest perceived costs.

II. AIRPORT GROUND ACCESS AND EGRESS MODEL

The model represents the passenger transport to and from the airport. The access part of the model represents the transport from the point of origin, which could be at home or at office, to the airport departure hall entrance from where the Airport Terminal model (PaxMod) begins. The egress part of the model represents passenger transport from the airport arrival hall to the destination. The air passenger access and egress transport is connected with many activities. These mainly include the passenger's choice of transport mode, time planning (departure from the point of origin, the time reserve desired) and the actual transport to the airport. The modelling of passenger traffic from and to the airport depends on many factors from which the key ones are:

- Flight schedule
- Aircraft size and load factor
- Party size distribution
- Type of flight (scheduled/charter)
- Type of passenger (business/leisure)
- Passengers' spatial distribution within the airport catchment area
- Passenger's transport mode choice

These factors are integrated in the AGAP model. The process diagram of the model is shown on Fig. 3.

A. Flight Schedule

Flight schedule is the primary input to the AGAP model. It is the starting point for the model. Following algorithms within the AGAP model are using its data to generate passengers within the catchment area. The most important flight schedule data are the aircraft arrival and departure times, the aircraft capacity, the average load factor and whether the flight is scheduled or charter. Our flight schedule is based on CFMU data and the data from [13]. For the simulation purposes we used the data from the flight schedule valid on one representative day. The selected day was 8th July 2008, which was the busiest day in terms of passenger throughput at Bratislava airport in 2008. According to data that were provided by Operation Division of Bratislava airport, 49 arrivals and 45 departures of commercial passenger aircraft took place at Bratislava airport on 8th July 2008. These aircraft movements generated passenger flows of 5,497 departing and 5,900 arriving passengers, which passed through the terminal at Bratislava airport on that particular day.

B. Charter/scheduled party size profile

Party size profile is one of the parameters that describe the passenger behaviour. This parameter describes the groups of passengers travelling together. The most common groups in this sense are couples, families, friends or colleagues. There are significant differences in party size distribution considering the scheduled flights and charter flights. Data regarding party size shown in Tab 1 and Tab 2 were gathered from [16].

TABLE I. PARTY SIZE PROFILES FOR LEISURE PASSENGERS
AT BRATISLAVA AIRPORT

Party Size	Count	Percentage
1	2095	49.45%
2	1523	35.95%
3	313	7.39%
4 and more	306	7.22%
TOTAL	4237	

TABLE II. PARTY SIZE PROFILES FOR BUSINESS PASSENGERS
AT BRATISLAVA AIRPORT

Party Size	Count	Percentage
1	2570	65.83%
2	971	24.87%
3 and more	363	9.30%
TOTAL	3904	

C. Allocation of passenger groups to the flight

In the process of allocation of passenger groups to the flight based on party size distributions the model randomly generates groups of passengers and fills the aircraft taking into account the seat capacity and the load factor. The random generation of the groups is designed as follows. From the party size profile the percentage of occurrence of each group is put into the chart in a cumulative way as it is depicted on the Fig. 1. Random percentage is generated according to the uniform distribution. This number is found on the vertical axis and from that point horizontal line is drawn against the group bars. Depending on which group bar the line crosses the group is selected. In the example on Fig. 1 there are two numbers generated 40% and 98%. According to the chart the number 40 transforms into the single passenger group and the number 98 transforms to the three or more passengers group. This generation of the groups goes in the cycle and the passengers are cumulated in the aircraft. Once the number of passengers reaches the aircraft capacity multiplied by load factor the group generating algorithm stops.

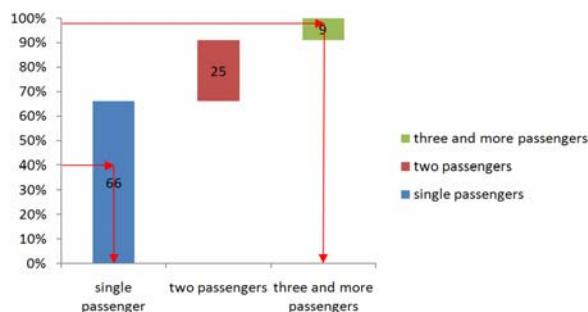


Figure 1. Random generator of passenger group size

D. Allocation of passenger groups to particular regions and cities

To be able to generate landside passenger trips to and from the airport it is necessary to know where the passengers start and end their trips. This can be derived from the passenger demand distribution within the airport catchment area. Air passenger demand distribution related data were gathered from the database of passenger questionnaire responses that was provided by [16]. It provides information about the demand distribution of various passenger groups within the country. However the distribution is based on eight autonomous regions of Slovakia and it is not subdivided further. To be able to generate passenger trips down to the cities we accepted following assumptions. All passengers within one group are assumed to be travelling together to/from the same city. The passenger demand within one single autonomous region in Slovakia is assumed to be uniformly distributed. Based on these assumptions and the data provided, we designed algorithm that allocates the city for each passenger group. The probability of allocation of the passenger group to the city is proportional to its population. Like this the algorithm firstly allocates the region to the passenger group based on the survey data and secondly allocates the city to the group based on the population distribution in the cities within the region.

E. Allocation of transport mode to charter/schedule groups

The process of allocation of the transport mode to the charter or schedule group is based on passenger's choice among available transport modes. In our model we selected following representative transport modes:

- ‘Kiss and drive’: (Passenger is driven by car to the airport by someone else)
- ‘Park and fly’ (Passenger drives and parks the car at the airport)
- Taxi
- Public transport – combination of trains and busses

In the model the transport mode choice is based on the evaluation of transport costs while choosing the transport mode with the lower perceived costs. The perceived costs of transport consist of the financial costs, the costs of time and transfer costs. The financial costs represent the money value necessary

to get from the place of origin to the airport and back including all related fees for example parking fees in case of car transport. The time costs represent the total travel time multiplied by the value of passenger travel time. The transfer costs represent a perceived value of additional physical and cognitive effort resulting from the transfer, and perceived value of risk of missing the connection.

III. AIRPORT TERMINAL PASSENGER FLOW MODEL - PAXMOD

The airport terminal passenger flow model (PaxMod) represents air passenger activities at the airport that start at entering the airport terminal and end after boarding an airplane. The flow input to the PaxMod is the flow generated by AGAP model. There are many activities that passenger does in airport terminal. These include visiting restaurants, the shopping, the renting a car etc. For the purposes of our research we are focusing only on activities related with the flight. These activities are divided into passenger processes and passenger movements. Passenger processes are mainly check-in, passport control, security check, customs, gate check-in and baggage claim. Passenger movements represent passenger walking from one service to another (e.g. from check-in to security).

A. Processes

In our literature review we identified three modelling approaches to model processes. These were stochastic queuing models, deterministic queuing models and simulation models. For the modelling of the processes we chose deterministic approach based on the work done by [14] and by [7]. The main reason for this is that we are interested in the flow from global view rather than from the view of individual passenger. Individual characteristics and microscopic level of modelling could be realised in microscopic simulation model. However the more complex the system is the more the simulation model tends to be difficult to develop. On the other hand application of queuing theory in stochastic queuing models removes some complexity as it is in the simulation models; however it is often based on strong assumptions which tend to be unrealistic. As an example queuing models hardly can capture varying rate of arrivals to the system which often occurs at the check-in counters at airports [11]. The deterministic approach allows modelling any kind of arrival profile and still the model could be relatively simple to develop so it might cause fewer difficulties in its development phase than in the case of the microscopic simulation model. Lastly the building blocks of our model should be transparent. Therefore we used relatively macroscopic level of modelling whereas only the behaviour of a group of passengers is modelled and not the individual behaviour.

The modelling approach is based on that the cumulative number of arriving passengers to the server (arrival profile) and the cumulative number of departing passengers from the server (departure profile) is known. It could be represented by $A(t)$ and $D(t)$ functions for arrival and departure profile respectively as it is depicted on Fig. 2

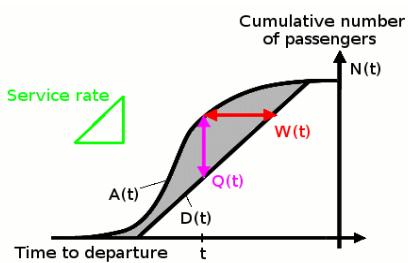


Figure 2. Cumulative diagram of passenger arrivals and departures from a server

From these functions average waiting time could be calculated as follows. Every passenger waits in the line certain time ranging from zero to some value. Sum of all waiting times could be calculated as an area bounded between $A(t)$ and $D(t)$ function:

$$T_{\text{wait}} = \int (A(t) - D(t)) dt$$

The cumulative number of passengers at the time t is represented by $N(t)$. Thus average waiting time per passenger until the time t is:

$$t_{\text{wait_avg}} = \frac{T_{\text{wait}}}{N(t)} = \frac{\int (A(t) - D(t)) dt}{N(t)}$$

B. Movements

Movements in PaxMod represent passenger walking from one server to another e.g. walking from check-in to the security control. The movements are modelled by shifting the departure profile from the server by specific time delay. The time delay is a time needed for the passenger to get from one server to another. Due to simplicity it is assumed that all passengers get to subsequent server within same period of time. Each subfunction of the departure profile is shifted by the same time delay. If universal form of polynomial of 3rd degree is written as:

$$P(t) = a_3 \cdot t^3 + a_2 \cdot t^2 + a_1 \cdot t + a_0$$

then the shifted function by the time delay d has following form:

$$\begin{aligned} P(t - d) &= a_3 t^3 + (-3a_3 d + a_2)t^2 \\ &\quad + (3a_3 d^2 - 2a_2 d + a_1)t \\ &\quad - a_3 d^3 + a_2 d^2 - a_1 d + a_0 \end{aligned}$$

C. Initial arrival profile

The initial servers of the PaxMod airport terminal model are check-in desks. Arrival profile to the check-in desks are based on arrival earliness profile gained from AGAP model. PaxMod is based on polynomial functions representing cumulative passenger arrivals, service and departures. AGAP model provides cumulative arrivals in a microscopic form. It means that each passenger arrival is represented by a time stamp and that is stored in a table in a cumulative form.

To feed the AGAP arrival earliness profile to the PaxMod it is necessary to represent AGAP profile with a polynomial function. My literature review showed that the polynomial functions of third or fourth degree are used. Within the PaxMod model the functions are further processed, combined and other data are from them calculated. Polynomial functions of fourth and higher degree are very complicated to process further. Therefore in PaxMod model the polynomials of third degree are used to represent passenger cumulative arrivals and departures. To fit the polynomial of third degree to the AGAP arrival earliness profile the linear regression is used.

D. Simulation and results

The Door-to-Gate Air Passenger Flow Model is used to simulate two scenarios of airport configuration - the baseline scenario and Airside-Landside separated scenario. The baseline scenario represents the classic concept of air passenger transport. The passenger leaves from home or work, travels by the public transport or by car to the airport and proceeds through the airport facilities to the aircraft. The Airside-Landside separated scenario (ASLS scenario) represents new concept of air passenger flows. This scenario is compared with the baseline scenario. The principal difference in the ASLS scenario is that passengers start the terminal processes in the hypothetical city-air-terminal collocated with City main railway station. In the ASLS scenario the passenger processes are different than those in Baseline scenario in following ways:

- The passengers are transported to the airport using hypothetical dedicated train.
- The check-in service, border control and the security are scheduled analogical way as in the Baseline scenario but are shifted by the transport time in advance.
- The check-in, border control and the security are operating in the appropriately equipped hypothetical train so that the passenger may be processed during the transport to the airport.

The results of the simulations are shown in Tab 3. It was shown that the ASLS concept performs worse for most of the passengers in terms of door-to-gate transit time. This is especially the case of passengers travelling by car. Passengers that start their journey outside of Bratislava and travel with public transport, spend approximately equal time in both concepts.

TABLE III. DOOR-TO-GATE TRANSIT TIMES [HH:MM]

Transport mode	Starting point	Classic	ASLS	Diff. ASLS-Classic
Car	Outside Bratislava	3:54	4:42	0:48
Public transport	Outside Bratislava	6:39	6:41	0:02
Car	Bratislava	1:55	2:18	0:23
Public transport	Bratislava	2:25	2:54	0:29

IV. CONCLUSION

For the evaluation of future airport concept from passenger flow perspective the door-to-gate air passenger flow model was presented. The model is based on airport ground access and egress passenger flow generator that uses random number generation based on probabilistic distributions and on airport terminal passenger flow model that uses deterministic queuing models for flow representation. Preliminary simulation results of passenger flows through selected airport concept called Airside-Landside Separation Concept showed that the concept has negative impact on passenger travel time in general.

REFERENCES

- [1] Marc Brochard. The airport of the future or breaking the constraints between the terminal and the runways. In Innovative Research Activity Report 2004, pages 37–45. Eurocontrol Experimental Centre, 2004.
- [2] Martin Matas. Future airport concept. In Activity Report 2005, pages 83–92. Eurocontrol Experimental Centre, 2005.
- [3] N. Ashford, N. Hawkins, and M. O’Leary. Passenger behavior and design of airport terminals. *Transportation Research Board Record*, 588:19–26, 1976.
- [4] G. F. Newell. Application of queuing theory. Chapman and Hall, 1971.
- [5] R. de Neufville and A. Odoni. Airport systems planning design and management, pages 134–135. McGraw-Hill, 2003.
- [6] Robert Horonjeff. Analyses of passenger and baggage flows in airport terminal building. *Journal of Aircraft*, 6(5):446–451, 1969.
- [7] Lorenzo Brunetta, Luca Righi, and Giovanni Andreatta. An operations research model for the evaluation of an airport terminal: Slam(simple landside aggregate model). *Journal of Air Transport Management*, 5:161–175, 1999.
- [8] M.R. Gatersleben and S.W. van der Weij. Analysis and simulation of passenger flows in an airport terminal. In Proceedings of the 1999 Winter Simulation Conference, pages 1226–1231, 1999.
- [9] Kiran A. S., Cetinkaya T., and Og S. Simulation modeling and analysis of a new international terminal. In Proceedings of the 2000 Winter Simulation Conference, pages 1168–1172, 2000.
- [10] Babeliowsky M. Designing interorganizational logistic networks: A simulation based interdisciplinary approach. PhD thesis, Technische Universiteit Delft, 1997.
- [11] P.E. Joustra and N.M. van Dijk. Simulation of check-in at airports. In Proceeding of 2001 Winter simulation conference, pages 1023–1028, 2001.
- [12] Winnie Daamen. Modelling Passenger Flows In Public Transport Facilities. PhD thesis, Technische Universiteit Delft, 2004.
- [13] Flight timetable Bratislava Airport Summer 2007
- [14] G. F. Newell. Application of queuing theory. Chapman and Hall, 1971.
- [15] Milan Stefanik. Problems of Airport Capacity Assessment, Doctoral Thesis, Žilinská univerzita v Žiline, 2009.
- [16] Airport Bratislava, Passenger surveys conducted by Marketing and Commerce Division of Bratislava airport in summer months of the years 2003, 2004 and 2007.

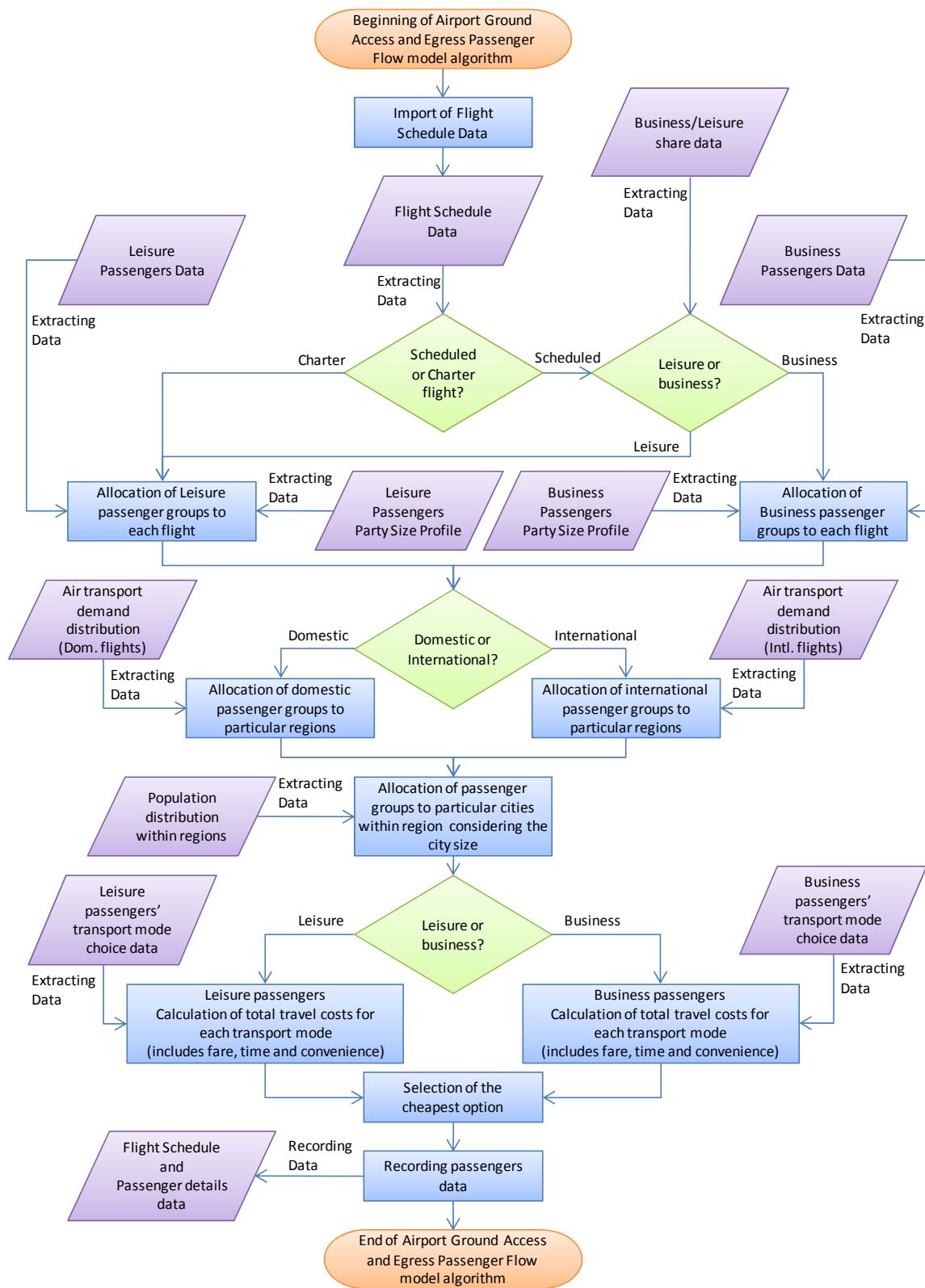


Figure 3. Airport ground access and egress passenger flow conceptual model [15]

The Airport Ground Movement Problem: Past and Current Research and Future Directions

Jason A. D. Atkin, Edmund K. Burke, Stefan Ravizza*
 School of Computer Science
 University of Nottingham, Jubilee Campus
 Nottingham, NG8 1BB, UK
 Email: {jaa,ekb,smr}@cs.nott.ac.uk

Abstract—Determining efficient airport operations is an important and critical problem for airports, airlines, passengers and other stakeholders. Moreover, it is likely to become even more so given the traffic increases which are expected over the next few years. The ground movement problem forms the link between other airside problems, such as arrival sequencing, departure sequencing and gate/stand allocation. This paper provides an overview, categorisation and critical examination of the previous research for ground movement and highlights various important open areas of research. Of particular importance is the question of the integration of various airport operations and their relationships which are considered in this paper.

Index Terms—Airsides airport operations, ground movement, taxiing, survey, future work, integration of airport operations.

I. INTRODUCTION

There has been a significant increase in air traffic over the past few years and this trend is predicted to continue. The SESAR (Single European Sky ATM Research) project predicts a doubling in the number of flights between 2005 and 2020 [1]. The project aims to triple capacity by 2020 and to reduce delays on the ground and in the air [2]. It is apparent that the hub airports often form bottlenecks for the overall air traffic management system within Europe. Hence, improvements in critical airport operations will be more and more important in the near future. The main operations which affect this bottleneck are arrival and departure management (sequencing and scheduling) at the runway [3]–[7], gate assignment [8], and ground movement.

The majority of the existing research has focussed on the optimisation of a single airport operation at a time. However, from both an economic point of view (reducing delays and increasing throughput), and an environmental point of view (reducing noise, air pollution and carbon emissions), there are obvious benefits to be gained from treating the different airport operations as a whole.

Ground movement links the various other operations together, and is the focus of this paper which provides, for the first time, a survey and comparison of the existing optimisation approaches within this field. Our purpose is to pinpoint the important open areas, of which, integrating the different airport operations is perhaps the most important potential future research direction.

*Corresponding author.

The remainder of this paper is structured as follows: Section II provides a description of the airport ground movement problem and relates it to the other relevant airport operations. Next, the existing models and solution approaches are discussed and categorised in Section III. We then highlight various important future research directions in Section IV, before ending the paper in Section V with some conclusions.

II. PROBLEM DESCRIPTION

The airport ground movement problem is basically a routing and scheduling problem. It involves directing aircraft to their destinations in a timely manner, with the aim being to either reduce the overall travel time and/or to meet some target time windows. Throughout the movement, it is crucial for reasons of safety, that two aircraft never conflict with each other. The complexity of the problem can vary and should drive the choice of solution approach. When an airport has only a few aircraft moving at once, with few potential conflicts between them, optimal routing can be achieved by simply applying a shortest path algorithm, such as Dijkstra's algorithm [9], [10], to each aircraft in turn. For larger airports, especially during peak hours, the interaction between the routes of different aircraft often requires the application of a more complex simultaneous routing algorithm.

The details of the problem descriptions and the constraints which have been utilised in previous work have varied according to the requirements of the airport which was being modelled. The various constraints upon the ground movement problem are considered in Section II-A. Since it is important for improving the operations at an airport to integrate the related operations with the ground movement problem, this integration is discussed in Section II-B, after which, the different objectives are described in Section II-C.

A. Constraints

The different constraints upon the problems discussed in the existing ground movement research literature can be divided into the following categories:

1) *Consideration of the route taken*: It is important to ensure that aircraft follow a permitted route. If the route for each aircraft is pre-determined, the ground movement problem is reduced to finding the best possible schedule [11], [12]. The other extreme occurs when no restrictions are set for the

routing of each aircraft [13]–[16]. The last possibility is for the restrictions to lie somewhere in between these extremes, where there is a predefined set of routes for each aircraft and the algorithm can choose amongst them [17]–[26].

2) *Separation constraints between aircraft*: As previously mentioned, it is crucial that aircraft do not conflict with each other and have a separation based on jet blast. This is ensured during taxiing by applying separation constraints. The required minimum distances between aircraft appear to vary between authors. For example, Petic et al. required it to be at least 60 metres [17], while Smeltink et al. required a value of 200 metres [11]. Such constraints can also depend upon the aircraft type or size. If an aircraft is at a gate, no such restriction is usually used. At the point of take-off or landing, other restrictions are employed, which are presented in Section II-B.

3) *Aircraft movement speeds*: Different aircraft require different lengths of time for taxiing. Recent research has taken this into account, modelling the speed depending either upon the type or size of an aircraft [23], [24], or the kind of taxiway that is being followed [18]. The time for making a turn can also be taken into account [17].

4) *Timing constraints for arrivals*: Arriving aircraft have to be routed from the runway to their stands. From the point of view of the isolated ground movement problem, the arrival time for aircraft can be considered to either be fixed or to permit small deviations. The allocated gate is usually assumed to be vacant and, therefore, the aim is usually for the aircraft to reach the gate as soon as possible, since this is better from an environmental as well as an airline and passenger perspective.

5) *Timing constraints for departures*: Departing aircraft have to be routed and scheduled from their stands to the runway from which they will be departing. A pushback time (or earliest pushback time) is usually provided and is often seen as an earliest time for an aircraft to start taxiing. The aims for the ground movement of the departing aircraft can be more complicated than for arrivals. Assuming that the departure sequencing has not been integrated into the problem, one of the following aims is usually adopted: 1) To reach the runway as early as possible. 2) To reach the runway in time to attain, or be as close as possible to, a pre-determined take-off time. 3) To reach the runway in time to take off within a specified time window, since many European aircraft have fifteen minute slots which are allocated by the Eurocontrol Central Flow Management Unit (CFMU) and have to be satisfied [20].

B. Integration of other airport operations

The ground movement problem does not actually occur in isolation at an airport. The arrival sequence will determine the times at which some aircraft enter the system, the gate/stand allocation problem will determine where they leave the system and where departures enter the system. The departure sequencing problem determines the times at which departures leave the system. These systems can be seen to be intimately linked, so potential benefits from integrating all four problems are obvious. However, little research so far has considered this integration. The complexity of these problems is such that it

is currently impossible to simultaneously optimise all of these airport operations, but the real situation at the airport means that there has to be at least some coordination between the solutions of the sub-problems.

1) *Integration of departure sequences*: For departing aircraft, the ground movement can affect the departure sequencing, and vice versa. An optimal take-off sequence is of no use if it cannot be achieved by the taxiing aircraft, as discussed in [6]. To maximise the throughput of a runway, two sequence-dependent separations are of major importance [27]: wake vortex separations and en-route separations. The wake vortex separations depend upon the weight classes of the aircraft, so that larger separations are required whenever a lighter class of aircraft follows a heavier class. Separations also have to be increased when aircraft have similar departure routes (to ensure that en-route separations are met) or when the following aircraft is faster (to allow for convergence in the air).

Departure sequencing is sometimes considered within ground movement research [18], especially the newer research [12], [15], [16], [25], [26], in order to ensure that aircraft arrive at the departure runway at appropriate times, rather than merely reducing the overall taxi times. Only wake vortex separations are usually considered. However, the en-route separations are also sometimes taken into account [15], [16].

Similarly, taxi times cannot be ignored in realistic departure sequencing systems. The movement near the runway is especially important, for example, within flexible holding areas [3], [6], or the interleaving of runway queues [28]. Even where the models for movement are not explicitly required, accurate taxi time predictions are often beneficial for improving sequencing [29], even when re-sequencing is performed at the runway, and would be even more important if the re-sequencing was performed earlier.

2) *Integration of arrival sequences*: Aircraft enter the ground movement system by landing on a runway, or by leaving stands. The entry times into the system of landing aircraft will influence the ground movement operations. Better arrival time predictions can have a positive effect on the ground movement planning. There may be a choice of landing runway to be made. This choice can depend upon the current status of the ground movement and the assigned gate for the aircraft. After landing it will influence the later ground movement planning.

In some airport layouts, runway crossings may be necessary for taxiing aircraft. For realistic runway sequencing and taxiing optimisation, such crossings may need to be taken into account [4], requiring knowledge of the runway sequencing when planning the ground movement. Furthermore, runways are sometimes used in mixed mode, in which case departure and arrival sequences also have to be coordinated [5], [7].

3) *Integration of gate assignment*: Gate assignment is another major problem which arises at congested airports. The aim is to find an assignment of aircraft to gates at terminals, or stands on the apron, so that some measure of quality, such as total passenger walking distance, is improved. This problem was fully discussed in a recent survey paper

by Dorndorf et al. [8], where the need for future work in multi-objective optimisation and robust assignments was also identified. The ground movement problem could be integrated with the gate assignment problem, with the aim being to allocate gates/stands so that the total taxiing distance is reduced. This would have a beneficial impact upon the use of fuel, with consequent benefits for the environment as well as financial savings for airlines, delay benefits for passengers and a reduction in congestion on the apron.

C. Objective functions

The aim of the ground movement problem depends upon the scope of the problem. Much of the previous research has concentrated upon minimising the total taxi time including the waiting time for aircraft at the runway [12], [13], [17], [24], while other research has considered makespan (the duration from first to last movement) minimisation [21], [22]. Yet more research has treated this as a multi-objective problem. For example, penalising deviations from a scheduled time of departure/arrival (STD/STA) [11], [23], [25], [26], or from the CFMU slots [20], in addition to considering one of the total taxi time or makespan reduction objectives. In other research, longer taxi paths were penalised as well [15], [16], [18]. Marín and Codina [14] used a weighted linear objective function to simultaneously consider the total routing time, number of controller interventions, worst routing time, delays for arriving and departing aircraft and the number of arrivals and take-offs.

D. Related research areas

Similar problems have been considered in other areas of research, such as the control of Automated Guided Vehicles (AGVs) [30], job-shop scheduling with blocking [31], train routing and scheduling [32] and airport surface conflict detection and resolution [33]. Of course, the details of the constraints and objectives differ, so there are limits to the applicability of the research.

III. EXISTING MODELS AND SOLUTION APPROACHES

In this section, we present a comparison and categorisation of the existing research for the ground movement problem at airports, which has previously taken two forms. The first form has involved the development of a Mixed Integer Linear Programming (MILP) formulation, to which a commercial solver was usually applied, yielding an optimal solution. Where models were formulated in a manner which would not be tractable to a MILP solver within a reasonable solution time, heuristic methods have been applied. This alternative approach has so far exclusively involved the use of Genetic Algorithms (GAs). Of course, as heuristics, GAs give no guarantee of the optimality of the solutions found. However, their success over far shorter (and far more realistic in practice) execution times can sometimes more than compensate for this.

We will first focus on the MILP formulations before discussing the GA-based approaches. For each approach, we will first discuss the various models which have been developed, before considering the previous research which has used these

TABLE I
OVERVIEW OF APPROACHES FOR THE GROUND MOVEMENT PROBLEM

Authors	Year	Approach	Representation
Pesic et al. [17]	2001	GA	Times
Gotteland et al. [18], [19]	2001/3	GA	Ordering, Times
Gotteland et al. [20]	2003	GA	Ordering
Smeltink et al. [11]	2004	MILP	Ordering
García et al. [21], [22]	2005	GA	Times
Marín [13]	2006	MILP	Times
Balakrishnan and Jung [23]	2007	MILP	Times
Marín and Codina [14]	2008	MILP	Times
Roling and Visser [24]	2008	MILP	Times
Deau et al. [25], [26]	2008/9	GA	Ordering
Keith and Richards [15]	2008	MILP	Ordering
Rathinam et al. [12]	2008	MILP	Ordering
Clare and Richards [16]	2009	MILP	Ordering

models in more depth. We will then compare the approaches, discussing the advantages and disadvantages of each. Finally, we end this section by considering two important issues: firstly, how do the models handle the dynamic nature of the real problems at the airports, and secondly, how can speed uncertainty be handled to make the solution more robust in the real situation? An overview of the published ground movement optimisation research considered here can be found in Table I, showing in chronological order both the solution approach which has been adopted and the defining characteristics of the model.

A. Mixed integer linear programming (MILP) formulations

MILP formulations are widely used by exact solution methods in operational research. In comparison to Linear Programming (LP) formulations where the objective function and constraints all have to be linear, MILP formulations introduce an additional restriction of integrality for some variables. Unfortunately, since this restriction changes the nature of the search space from continuous to discrete, it often leads to problems which are much harder to solve, so that solution times for large problems may no longer be practical.

Three different MILP modelling approaches, which have been adopted, are described below:

- Exact position approach: Here a time is allocated for each aircraft to traverse each individual part of its path. The approaches of Marín [13], Balakrishnan and Jung [23], Marín and Codina [14] and Roling and Visser [24] used a space-time network for this purpose. A spacial network representing the map of the airport is used as a starting point, then time is discretised and a copy of the underlying spacial network is created for each time unit. These are then used to build a time expanded network. A good illustration of this can be found in Marín and Codina [14].
- Ordering approach: In this case, rather than dealing directly with timings, the algorithm first aims to decide upon the sequencing, then uses this information to schedule times for each aircraft at each node or edge.

This approach was adopted by Smeltink et al. [11], Rathinam et al. [12], Keith and Richards [15] and Clare and Richards [16]. All of these only required a spacial network and modelled the sequencing constraints using binary variables, where the variables for a pair (i,j) of aircraft at a node/edge are equal to one if and only if aircraft i passes this node/edge before aircraft j . With this approach, the times for each aircraft can be modelled as continuous variables, avoiding the disadvantages of time discretisation.

- Immediate predecessor/successor approach: It would also be possible to indicate only the immediate predecessor and successor for each aircraft at each node/edge rather than a full sequencing. As far as we can determine, this approach has not been used for solving the ground movement problem so far. Although the model in Smeltink et al. [11] indicated the immediate predecessor aircraft, this was only to support the ordering model.

B. Review of previous MILP-related research

To our knowledge, Smeltink et al. [11] was the first approach to handle the ground movement problem using the MILP formulation. This was performed for Amsterdam Schiphol Airport in 2004. Since this airport used standard, predefined taxi routes for aircraft, the problem was reduced to a scheduling problem. The approach worked on a spacial network where times were modelled as continuous variables and binary variables were used for the sequencing, as described above. The objective was to minimise the waiting time while taxiing and the deviation between the desired departure time and the scheduled departure time.

In 2006, Marín [13] presented a linear multi-commodity flow network model to simultaneously solve the aircraft routing and scheduling problem around airports. Two different methodologies were used to solve the MILP formulation: a branch and bound, and a fix and relax approach. In the latter case, the planning period was split into k smaller periods. Initially, only the variables within the first time period are taken as binary and a linear relaxation is applied to the variables for the other periods. The variables for the first period are then fixed, the variables for the second time period are made binary and the linear relaxation is maintained for the remaining variables. This is repeated for all k periods until all of the variables have been fixed. The objective of the MILP formulation was to minimise the total taxi time.

Marín and Codina later published further work [14] where the model was multi-objective. The weighted linear objective function considered five other objectives, in addition to the previous goal of reducing the total routing time: 1) reducing the number of controller interventions, 2) reducing the worst routing time, 3) reducing the delays for arrivals, 4) reducing the delay for departures and 5) attempting to maximise the number of arrivals and take-offs. In contrast to other models, they allowed the aircraft to use the whole network and did not restrict them to a pre-determined set of paths. However, the presented algorithm was not able to deal with the separation

constraints in an accurate way because the constraints were only modelled in the space-time network, which is independent of the type or size of aircraft.

Balakrishnan and Jung [23] published another MILP formulation of the ground movement problem on a space-time network. In this approach, each aircraft could be allocated one of a limited set of routes. The relative benefits of different control approaches, such as controlled pushback and taxi path re-routing were also considered. Their aim was to minimise the total taxi time and to penalise situations where aircraft departed too late. It was pointed out that controlled pushback could reduce the average departure taxi time significantly, saving fuel.

An alternative MILP formulation for ground movement, which was also based on a space-time network, was provided by Roling and Visser [24]. A number of alternative routes were assigned to each aircraft beforehand, and only these were considered at the solution stage. It was possible for an aircraft to wait at the beginning of the journey, as well as on special nodes during the journey. The objective was to minimise a weighted combination of the total taxi time and total holding time at the gates. The objective function considered the entire route for each aircraft but the solution was only guaranteed to be conflict-free within the planning horizon, since these constraints were relaxed for later times.

Rathinam et al. [12] used a MILP formulation which was based on the work of Smeltink et al. [11] and primarily considered the ordering of the aircraft at nodes. Further separation constraints were added to the model, and it was simplified by reducing the number of binary variables. The algorithm used a spacial network and a predefined route for each aircraft, to minimise the total taxi time.

Keith and Richards [15] introduced a new model for the coupled problem of airport ground movement and runway scheduling. Their MILP optimisation was influenced by the work of both Smeltink et al. [11] and Marín [13]. The objective function was a weighted combination of minimising the makespan, the total taxi and waiting time and the total taxi distance. As in Smeltink et al. [11], a spacial network was used, with binary variables for handling the sequencing constraints and continuous variables for the timings. Although both wake vortex and en-route separations were considered for the take-off sequencing element, there were no route limitations applied. The work of Clare (nee Keith) and Richards [16] extended their previous work. Their MILP formulation was changed to make it possible to introduce an iterative solution method. In the first step, a relaxed MILP formulation was solved, and no guarantees were given for a conflict-free solution. An iterative procedure was then applied, where additional constraints were added where they were necessary to avoid any conflicts detected in the previous iteration. This was repeated until a conflict-free schedule was found.

C. Genetic algorithm (GA) models

GAs are search methods inspired by evolutionary biology. They incorporate the ideas of natural selection, mutation

and crossover [34]. GAs maintain a population of candidate solutions, have a method (called a fitness function) for evaluating solutions and apply a selection mechanism to guide the algorithm towards good solutions. The correct encoding of the problem can be key for the successful application of a GA (as we will consider in the next section), as can be the choice of appropriate mutation and crossover operators for the selected problem encoding.

We now consider the important elements of the encodings which have been used for the ground movement problem over the last decade before considering, in Section III-D, the specific encodings. As for the MILP approaches, the GAs consider either the absolute timing or the relative sequencing of the ground movement.

All of the encodings which have been considered in the GA implementations, [17]–[22], [25], [26], included the route allocation information, specifying the route r_i to allocate for each aircraft i . The additional information which was included differed between the approaches, but can be summarised into three categories:

- Applying an initial (aircraft-specific) delay/hold time, prior to pushback. The GA is responsible for determining this delay for each aircraft, as well as the route to allocate. This approach was adopted by [21], [22].
- Applying a delay at some point during the movement, and not restricting it to being applied at the start of the taxiing. This could be implemented either by specifying times for both initiating and terminating the delay (the approach which was adopted in [17], [19]) or as a delay amount and (spacial) position at which to apply it to the aircraft, as in [18]. The GA is responsible for investigating when or where to apply the delay and the duration or end time of the delay as well as the route to allocate to the aircraft.
- Prioritising aircraft movement, where the GA is used to investigate the relative prioritisation of the aircraft rather than allocating holds directly. Here, the priority determines which aircraft take precedence when there are conflicts during the movement. This approach was adopted in [18]–[20], [26], where the GA investigated the priorities to assign to aircraft as well as the routes.

D. Review of previous GA-related research

As far as we can determine, Pesic et al. [17] published the first paper for optimising the ground movement problem at airports in 2001. They allowed a single delay per aircraft at a time determined by the GA. Their fitness function considered the number of time steps C , for which aircraft were in conflict during the movement, and the total travel time T for aircraft. The GA aimed to maximise the fitness value, which was $\frac{1}{2+C}$ in the presence of conflicts or $\frac{1}{2} + \frac{1}{T}$ in the absence of conflicts. All values bigger than $\frac{1}{2}$ corresponded to solutions which were conflict-free and all values smaller than $\frac{1}{2}$ had at least one conflict and were therefore infeasible. Crossover and mutation operators were introduced along with a diversification strategy and some simple termination criteria. For a random pair of parent solutions, the crossover operator chose for each aircraft

the parent which had fewer conflicts with other aircraft, in order to increase the probability of producing an offspring population with better fitness values. This operator was appropriate because the problem was partially separable [35]. The mutation modified the details for the aircraft with the (potentially shared) worst local fitness value.

Gotteland et al. [18] extended their previous work by considering how the GA could deal with speed uncertainty. We believe that this is an important consideration and will discuss it in Section III-G. In addition to the encoding from their previous work [17], they used a representation for prioritising aircraft movements, discussed in Section III-C. The encoding included the route number and priority level for each aircraft. A fitness value was computed by applying an A* algorithm with the specified prioritisation of the aircraft. A space-time network was then generated and aircraft were routed in order of priority level. After an aircraft had been routed, the network was adjusted in such a way that the allocated route was removed, along with all potentially conflicting edges, so that the routing of the next aircraft avoided conflicts with previous aircraft.

The clustering of aircraft within these ground movement problems was considered in [18]. A two stage approach was adopted, where the clusters of aircraft with conflicts were solved independently in the first stage, before the different clusters were unified and solved in combination in the second stage.

Gotteland et al. [19] subsequently presented an alternative sequential algorithm: a branch and bound algorithm, with a first search strategy replacing the A* algorithm to speed up the calculation of the fitness value, since there is always a preference to continue taxiing rather than to hold position.

Gotteland et al. [20] explained the way in which their GA handles both take-off time prediction and CFMU slots. They modified their algorithms from [18] with the aim of reducing the deviation from CFMU slots (rather than minimising the necessary taxiing time) by penalising (with a linear cost) deviations from the desired take-off times for each aircraft, with a steeper penalty when the scheduled take-off is outside the CFMU slot.

García et al. [22] hybridised two earlier approaches which were previously detailed by the same authors in [21]. A modified minimum cost maximum flow algorithm determined the initial population of a GA and was used to penalise the fitness function. The approach considered the application of an initial delay at the gate and the allocation of a route to each departing aircraft, with no possibility for waiting at intermediate points or slower taxiing during the ground movement. They used tournament selection, single-point crossover, a traditional mutation operator and an additional random variation of the delay time. Their fitness function penalised infeasible solutions and tried to minimise the makespan and the sum of the delays, while attempting to maximise the number of departing aircraft.

Two more recent papers from Deau et al. [25], [26], developed the ideas which have been discussed for [17]–[20]. They proposed a two-phase approach which considered

the runway sequencing in the first stage and the ground movement in the second stage. The separations to account for the wake vortices were the most important constraint for the runway sequencing element. A deterministic constraint satisfaction problem solution algorithm was used, which was based on a branch and bound methodology. They used an objective function which was similar to that which was used in Gotteland et al. [20]. Departing aircraft were moderately penalised if their scheduled time deviated from the desired time within the CFMU slot, but were much more heavily penalised if the scheduled time was outside this slot. Arriving aircraft had a fixed predicted time to land, so a solution was only feasible if these aircraft had, at most, a small delay (no more than one minute) compared with the predicted landing time. In the second stage, their GA was modified to find a good solution for the ground movement problem given the runway sequencing from the first stage. The target runway sequence was considered as the ideal result of the routing stage, but was not treated as a hard constraint, thus, the fitness function for their GA penalised deviations from the target times.

E. Comparison of the approaches

We now consider the major differences between the different models and solution approaches.

1) Differences in objectives and constraints: The optimisation of airport operations is a real-world problem, and as such it is important that the real objectives of the airport and real constraints upon the problem are considered. The majority of the published work has considered real airport settings, and it is apparent that both the objectives and the details of the constraints have differed between airports. Consequently, the models for the problems have also differed, resulting in the development of different solution approaches.

2) Optimality vs. execution time: The solution approach which is adopted may also depend upon the load upon the airport (i.e. the number of aircraft which need to be simultaneously considered), since exact solution approaches become less practical as loads increase. With the expected increases in the density of air traffic meaning that airports have to be able to handle more aircraft in the near future, some solution approaches may potentially need to be adjusted over time.

It is well known that GAs are heuristics rather than exact solution methods and can, therefore, often give neither any guarantee for the solution nor even an approximation ratio in many situations. However, a poor formulation of a MILP can also mean that an exact solution to the MILP can be a poor solution for the underlying real-world problem. For example, with time discretisation models, the way in which the time discretisation is handled can have a major effect upon the optimality of the results: smaller intervals may give better results but will result in significantly larger problems to solve. Similarly, the way in which a model deals with the separation rules between aircraft can affect the quality of the results. It should be noted that none of the papers which were discussed here measured the optimality gap for realistic scenarios, evaluating the effects of utilising only a heuristic

(GA-based) solution approach or of the effects of time discretisation, perhaps due to the difficulty or impracticality of optimally solving these problems. In our opinion, it would be worthwhile to have some kind of comparison between the performance of the approaches, to be able to see the trade-off explicitly.

Due to the fact that airports are usually interested in real time decisions, the execution time of an algorithm is a crucial measure. From this point of view, heuristics such as GAs outperform MILP formulations. For example, in [24] it was shown that the execution time increased dramatically as the number of aircraft increased.

Different researchers have also used different objective or fitness functions, due to having slightly different aims. We believe that the generation of some generic benchmark scenarios to allow such an analysis to be performed, comparing exact and heuristic solution approaches and the effects of different objective functions, would be of huge benefit and is a path down which we plan to proceed.

As far as we are aware, there has been no investigation using other metaheuristics such as simulated annealing [36], or tabu search [37]. Furthermore, there seems to be an unexploited potential for hybrid approaches which can make use of the advantages of different models.

F. Dealing with the dynamics

One major characteristic of the problem of ground movement at airports is the dynamic nature of the problem. Predictions become less accurate the further they are in the future: predicted positions for current aircraft may be wrong as may be predictions of when new aircraft will be ready to pushback from the gates or to land. Predictions, therefore, have to be regularly updated and, since some approaches need a significant execution time, attempts have been made to decompose the problems into smaller sub-problems. In this section, we summarise the approaches which have been used to cope with the dynamic nature of the routing problem.

- A simple modelling approach, by the name of *shifted windows*, was introduced by Petic et al. [17] for their GA. Every Δ minutes, the situation was resolved for a fixed time window. Only arriving or departing aircraft within the time window were considered but the time window was enlarged for these aircraft to avoid horizon effect problems.
- Smeltink et al. [11] evaluated three different variants of a *rolling horizon* approach, not only for handling the dynamics of the problem, but also to reduce the size of the problem to be solved. In each case, the planning period was split into disjoint, equal length time intervals. In the first variant, the routes which had been allocated in previous intervals were considered to be fixed, while in the second variant they could be modified. In the third variant, the aircraft were sorted according to their pushback or landing time, respectively, and a *sliding window* was applied to consider m aircraft in each iteration. The first iteration considered aircraft 1 to m , then aircraft

1 was fixed and aircraft 2 to $m + 1$ were considered, then aircraft 2 was fixed, and so on. Unfortunately, this variant had a significantly higher execution time without increasing the solution quality significantly.

- The fix and relax approach (discussed in Section III-B) which was used by Marín [13] for solving his MILP formulation, worked in a similar way to the sliding window approach. He also used an alternative time-interval-based approach, where only aircraft in a particular interval were used for planning but the interval was not enlarged to guarantee a conflict-free solution. Instead, a shortest path algorithm was used to estimate the remaining time for the aircraft which do not reach their destination within the interval.

G. Robustness and speed uncertainty

Almost all published approaches were based on deterministic data. However, the real world situation at airports is less predictable. Therefore, we think it is important to take solution robustness into consideration. Uncertainty in the data for the ground movement problem can appear in different areas, one of which is speed predictions. An approach to cope with this was presented and illustrated in Gotteland et al. [18]. They modelled the speed uncertainty as a fixed percentage of the predefined speed. Hence, an aircraft was assumed to occupy not only a single position in the network but multiple possible positions at the same time. While an aircraft was taxiing, the number of occupied positions grew and when an aircraft was waiting at a holding point, the speed uncertainty and number of occupied positions decreased.

IV. IMPORTANT FUTURE DIRECTIONS

In this section, we describe several important open research directions for the airport ground movement problem.

A. Consistency and comparability

As discussed in Section III-E, the constraints and objectives vary widely within the published research. No comparison has so far been performed between different approaches, so it is difficult to estimate the gap between the exact optimisation methods (e.g. MILP formulations) and the heuristic approaches (e.g. GA) for either the quality of the solution or the execution time of the algorithms. More consistency is desirable. For this reason, and in an attempt to promote research in this area, we have set up a repository for datasets for these problems¹ and intend to do some quantitative comparison.

B. Integration of other airport operations

The integration of other airport operations, such as departure and arrival sequencing and gate assignment, is highly desirable and, ultimately, optimisation across multiple airports would be even better. Of course, the complexity of the integrated problem would grow and, since the computation is time-critical, there seems to be more potential for heuristic and

¹Some datasets and details are available at <http://www.asap.cs.nott.ac.uk/atr/benchmarks/> and we encourage further contributions.

hybrid methods than exact approaches. With the integration of different airport operations, the problem may also have to be treated as a multi-objective optimisation problem.

C. Robustness and uncertainty

Uncertainty in the input data is common at airports. Push-back time uncertainty and taxi speed/duration uncertainty are known to be major limiting factors upon the accuracy of models. We see the need for more investigation into models of the airport ground movement problem which are more robust against such uncertainty.

D. Restricted stopping positions

It is easier to hold aircraft at some points (for example at lights built into the taxiways) than at others and, in some cases, it is reasonable to hold an aircraft in a specific position only under certain circumstances. For example, it is reasonable to ask a pilot to wait in a queue behind another aircraft, but may not be sensible to request a pilot to ‘taxi until 12:05 then pause for 30 seconds’. Different modelling and solution approaches can result in different operational modes. We suggest that the approach to adopt should be influenced by the real operating modes, so that the algorithmic results can correspond to instructions which could be given to pilots, ensuring that plans could actually be enacted.

E. Environmental considerations in taxiing

Consideration of the environmental effects of airports has become increasingly important and could be taken into account for ground movement. For example, where possible, delays for an aircraft should be scheduled prior to starting the engines, i.e. as initial delays at the gate/stand.

Perhaps more interestingly from the point of view of the problem modelling, aircraft engines are more efficient when a constant taxi speed can be maintained rather than having a lot of acceleration and deceleration. Speed changes and multiple stops should, therefore, be avoided or reduced. It may be advisable to consider some kind of post-processing to calculate speeds for link traversals, so that the pilots could be given appropriate information to allow them to replace higher speed taxi operations plus waits by a lower speed operation.

F. Limiting changes

When the real-world dynamic case is considered, it is possible that routes or sequencing can change over time. This may be highly undesirable if information has been transmitted to pilots. Thus, the effects of avoiding changes should at least be considered.

V. CONCLUSIONS

This work provides the first overview and comparison of the various ground movement models and solution methods in the literature. It is apparent that there are significant differences between both the objectives and the constraints which were utilised in previous research. To some degree this is inevitable due to the differences between airports and different stakeholder aims. However, there is obvious benefit

to be gained from a formalisation of these. The state-of-the-art approaches use either a MILP formulation or a genetic algorithm approach and a categorisation of the representations has been provided for both.

In addition to highlighting the state-of-the-art in this research area, a number of interesting and important future research directions have also been identified. Of particular importance is the integration of other (highly-related) airport operation problems. Runway sequencing (for both departures and arrivals) and gate assignment are highly connected to the problem of airport ground movement and we suggest that there would be benefits from handling them simultaneously. More consistency within airport operations would also be helpful and generic benchmark scenarios would be useful for both quantifying algorithms and encouraging further research by those who may not have direct contact with an airport. Finally, we have identified the importance of handling uncertainty in taxi speeds and generating robust solutions and of considering the operational limitations of communicating instructions to pilots and the environmental effects of decisions.

ACKNOWLEDGMENT

The authors wish to thank the Engineering and Physical Sciences Research Council (EPSRC) for providing the funding which made this research possible.

REFERENCES

- [1] SESAR, "Milestone deliverable D1 - air transport framework: The current situation," Eurocontrol, Tech. Rep. Edition 3, 2006. [Online]. Available: http://www.eurocontrol.int/sesar/public/standard_page/documentation.html
- [2] ———, "European air traffic management master plan," Eurocontrol, Tech. Rep. Edition 1, 2009. [Online]. Available: http://www.eurocontrol.int/sesar/public/standard_page/documentation.html
- [3] R. A. Leese, A. Craig, R. Ketzcer, S. D. Noble, K. Parrott, J. Preater, R. E. Wilson, and D. A. Wood, "The sequencing of aircraft departures," Study report from the 40th European Study Group with Industry, Tech. Rep., 2001. [Online]. Available: <http://www.smithinst.ac.uk/Projects/ESGI40/ESGI40-NATS/Report>
- [4] I. Anagnostakis and J.-P. Clarke, "Runway operations planning: A two-stage solution methodology," in *Proceedings of the 36th Annual Hawaii International Conference on System Sciences, Los Alamitos, USA*, 2003.
- [5] L. Bianco, P. DellOlmo, and S. Giordani, "Scheduling models for air traffic control in terminal areas," *Journal of Scheduling*, vol. 9, pp. 223–253, 2006.
- [6] J. A. D. Atkin, E. K. Burke, J. S. Greenwood, and D. Reeson, "Hybrid metaheuristics to aid runway scheduling at London Heathrow airport," *Transportation Science*, vol. 41, no. 1, pp. 90–106, 2007.
- [7] D. Böhme, R. Brucherseifer, and L. Christoffels, "Coordinated arrival departure management," in *Proceedings of the 7th USA/Europe Air Traffic Management R&D Seminar, Barcelona, Spain*, 2007.
- [8] U. Dorndorf, A. Drexl, Y. Nikulin, and E. Pesch, "Flight gate scheduling: State-of-the-art and recent developments," *Omega*, vol. 35, no. 3, pp. 326–334, 2007.
- [9] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, pp. 269–271, 1959.
- [10] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed. MIT Press and McGraw-Hill, 2001.
- [11] J. W. Smeltink, M. J. Soomer, P. R. de Waal, and R. D. van der Mei, "An optimisation model for airport taxi scheduling," in *Proceedings of the INFORMS Annual Meeting, Denver, USA*, 2004.
- [12] S. Rathinam, J. Montoya, and Y. Jung, "An optimization model for reducing aircraft taxi times at the Dallas Fort Worth International Airport," in *Proceedings of the 26th International Congress of the Aeronautical Sciences*, 2008.
- [13] Á. Marín, "Airport management: Taxi planning," *Annals of Operations Research*, vol. 143, no. 1, pp. 191–202, 2006.
- [14] Á. Marín and E. Codina, "Network design: Taxi planning," *Annals of Operations Research*, vol. 157, no. 1, pp. 135–151, 2008.
- [15] G. Keith, A. Richards, and S. Sharma, "Optimization of taxiway routing and runway scheduling," in *Proceedings of the AIAA Guidance, Navigation and Control Conference, Honolulu, Hawaii, USA*, 2008.
- [16] G. Clare, A. Richards, and S. Sharma, "Receding horizon, iterative optimization of taxiway routing and runway scheduling," in *Proceedings of the AIAA Guidance, Navigation and Control Conference, Chicago, USA*, 2009.
- [17] B. Pesic, N. Durand, and J.-M. Alliot, "Aircraft ground traffic optimisation using a genetic algorithm," in *Proceedings of the Genetic and Evolutionary Computation Conference, San Francisco, USA*, 2001.
- [18] J.-B. Gotteland, N. Durand, J.-M. Alliot, and E. Page, "Aircraft ground traffic optimization," in *Proceedings of the 4th USA/Europe Air Traffic Management R&D Seminar, Santa Fe, USA*, 2001.
- [19] J.-B. Gotteland and N. Durand, "Genetic algorithms applied to airport ground traffic optimization," in *Proceedings of the Congress on Evolutionary Computation, Canberra, Australia*, vol. 1, 2003.
- [20] J.-B. Gotteland, N. Durand, and J.-M. Alliot, "Handling CFMU slots in busy airports," in *Proceedings of the 5th USA/Europe Air Traffic Management R&D Seminar, Budapest, Hungary*, 2003.
- [21] J. G. Herrero, A. Berlanga, J. M. Molina, and J. R. Casar, "Methods for operations planning in airport decision support systems," *Applied Intelligence*, vol. 22, no. 3, pp. 183–206, 2005.
- [22] J. García, A. Berlanga, J. M. Molina, and J. R. Casar, "Optimization of airport ground operations integrating genetic and dynamic flow management algorithms," *AI Communications*, vol. 18, no. 2, pp. 143–164, 2005.
- [23] H. Balakrishnan and Y. Jung, "A framework for coordinated surface operations planning at Dallas-Fort Worth International Airport," in *Proceedings of the AIAA Guidance, Navigation, and Control Conference, Hilton Head, USA*, 2007.
- [24] P. C. Roling and H. G. Visser, "Optimal airport surface traffic planning using mixed-integer linear programming," *International Journal of Aerospace Engineering*, vol. 2008, no. 1, pp. 1–11, 2008.
- [25] R. Deau, J.-B. Gotteland, and N. Durand, "Runways sequences and ground traffic optimisation," in *Proceedings of the 3rd International Conference on Research in Air Transportation, Fairfax, USA*, 2008.
- [26] ———, "Airport surface management and runways scheduling," in *Proceedings of the 8th USA/Europe Air Traffic Management R&D Seminar, Napa, USA*, 2009.
- [27] J. Atkin, "On-line decision support for take-off runway scheduling at London Heathrow airport," PhD Thesis, The University of Nottingham, 2008.
- [28] M. A. Bolender, "Scheduling and control strategies for the departure problem in air traffic control," Ph.D. dissertation, University of Cincinnati, 2000.
- [29] J. A. Atkin, E. K. Burke, J. S. Greenwood, and D. Reeson, "On-line decision support for take-off runway scheduling with uncertain taxi times at London Heathrow airport," *Journal of Scheduling*, vol. 11, no. 5, pp. 323–346, 2008.
- [30] I. F. Vis, "Survey of research in the design and control of automated guided vehicle systems," *European Journal of Operational Research*, vol. 170, no. 3, pp. 677–709, 2006.
- [31] N. G. Hall and C. Sriskandarajah, "A survey of machine scheduling problems with blocking and no-wait in process," *Operations Research*, vol. 44, no. 3, pp. 510–525, 1996.
- [32] J.-F. Cordeau, P. Toth, and D. Vigo, "A survey of optimization models for train routing and scheduling," *Transportation Science*, vol. 32, no. 4, pp. 380–404, 1998.
- [33] J. García, J. A. Besada, G. de Miguel, and J. Portillo, "Data processing techniques for conflict detection on airport surface," in *Proceedings of the 5th USA/Europe ATM R&D Seminar, Budapest, Hungary*, 2003.
- [34] K. Sastry, D. Goldberg, and K. Graham, *Search Methodologies*. Springer, 2005, ch. Genetic Algorithms, pp. 97–125.
- [35] N. Durand and J.-M. Alliot, "Genetic crossover operator for partially separable functions," in *Proceedings of the third annual Genetic Programming Conference, USA*, 1998.
- [36] E. Aarts, J. Korst, and W. Michiels, *Search Methodologies*. Springer, 2005, ch. Simulated Annealing, pp. 187–210.
- [37] M. Gendreau and J.-Y. Potvin, *Search Methodologies*. Springer, 2005, ch. Tabu Search, pp. 165–186.

Modeling of Aircraft Surface Traffic Flow at Congested Airport Using Cellular Automata

Ryota Mori

Electronic Navigation Research Institute
Chofu, Japan
r-mori@enri.go.jp

Abstract— To manage the recent growth of air transportation, effective air traffic control and a 4-dimensional trajectory control concept have been already developed. However, most studies consider the flight phase only, which makes the airport surface congestion a bottleneck. The control of the airplane during the ground phase is almost entirely in the hands of the pilot and its uncertainty makes the simulation of airport traffic difficult. In addition, a congestion is a complicated phenomenon, not investigated in detail yet. This paper proposes a new airport surface simulation method considering a congestion phenomenon based on cellular automata. The floor field model is applied, and an aircraft speed decision process involving long-range interaction is developed. The effectiveness of this method is verified by comparing the results obtained with actual airport surface traffic data.

Keywords-component; *airport surface; cellular automata; airport simulation; NS model*

I. INTRODUCTION

The importance of air traffic management has been stressed on with the recent increase in air traffic. 4-dimensional (4D) trajectory control has become a keyword to the future air traffic control. Although recent aircraft can follow 3D trajectory very precisely, their arrival time is not estimated that accurately yet. The 4D trajectory concept considers time navigation, which aims at using airspace more effectively. However, the 4D trajectory is usually defined during the flight phase only, so the airport congestion becomes a bottleneck. Even if the aircraft can follow the 4D trajectory during the flight phase, this process becomes irrelevant if the aircraft cannot land at the airport on schedule. In addition, while many studies can be found about the 4D flight navigation, relatively small number of studies is observed about the airport traffic problem. The author believes that the airport congestion problem should be discussed in more detail, because considering airport surface traffic, the take-off and landing time of the aircraft can be estimated with high accuracy.

A straightforward approach could be a simulation of the airport surface traffic flow, if it were not for the obstacles stated below. First, the aircraft departure time is likely to be changed. If some passengers do not board the aircraft by the scheduled time, the aircraft has to wait for them. Second, there are many uncertainties in taxiing. The aircraft taxiing is mainly in the hands of the pilot, so the taxiing time varies from pilot to pilot. In addition, the aircraft taxiing is affected by other

aircraft taxiing. For example, if an aircraft blocks a taxiway, no other aircraft can proceed. If an aircraft goes slowly on the taxiway, other aircraft cannot overtake it. This paper focuses on the second problem, i.e., airport traffic congestion problem. Although there are some studies on the automation of airport surface traffic scheduling[1][2], the aircraft taxiing dynamics is not understood fully enough, which can be a critical factor in the overall improvement of air traffic management. Moreover, while there are some famous tools to support airport surface scheduling and management such as Surface Management System (SMS) which is used to improve efficiency of surface operation[3], the aircraft taxiing speed is assumed to be constant, which differs significantly from the actual airport surface traffic especially in the congested airport. Other papers propose the airport surface trajectory model where the taxiing speed is carefully considered based on the statistical data[4], but only unimpeded traffic is considered. The airport traffic problem becomes critical when an airport is congested. In this paper, taxiing dynamics of both impeded and unimpeded aircraft are simulated.

The congestion phenomenon is famous in the field of car driving, and many studies have been conducted to model it, e.g., optimal velocity model[5] and Nagel-Schreckenberg (NS) model[6]. In these models, traffic jam is simulated under the condition that all drivers follow the same driving rule. NS model is based on the cellular automata (CA), which makes simulation simpler. Although the basic NS model is applied for a single traffic lane, it is extended to multiple traffic lanes and crossings[7]. In the general CA model, whether a particle proceeds to the next cell or not depends only on this cell's availability, i.e. as long as the cell is empty, the particle moves forward exactly one cell. However, in the case of airport ground traffic, the flow is considerably smaller than that of highway traffic, so the application of the general CA model could not lead to meaningful results statistically. On the other hand, in NS model each particle is characterized by speed, too. That makes it possible for the particle to advance by more than a single cell, which turns the model from a pure random statistical one into a deterministic one. Therefore, it is considered that NS model can adequately describe the ground movement of aircraft at an airport. Despite the dynamic differences between aircraft and car traffics, the NS model can be successfully adopted after some changes which reflect the characteristics of ground aircraft movement.

This paper is organized as follows. Section II starts with brief overview of NS model, followed by an explanation of

certain airport traffic features which need to be considered explicitly when adopting NS model to ground air traffic. In section III, the problems are implemented in the proposed simulation model, which is discussed in detail. In section IV, using the actual airport surface data, a simulation model is constructed, and the simulation result is compared to the actual data. This paper is summarized in section V.

II. NS MODEL AND THE CHARACTERISTICS OF AIRPORT TRAFFIC

A. NS model

As the proposed model is based on NS model, firstly NS model is briefly explained. NS model is based on car traffic. The following explanation is about car driving. A simplified image of the model is shown in Fig. 1. A single lane road divided into cells of equal length is assumed. Each cell can contain a single car only, and each car is characterized by a non-negative integer velocity. The time is also discrete, where at each time step, each car moves its cell based on the velocity. All cars move in the same direction. The velocity is calculated based on the following rules. $v_i(t)$ and $x_i(t)$ indicate velocity and position of the i th car at time t , respectively. The cars are ordered from 1 st to n th.

a) Acceleration: $v_i(t) = \min(v_{\max}, v_i(t-1)+1)$

b) Crash avoidance:

$$v_i(t) = x_{i-1}(t) - x_i(t) \text{ if } x_{i-1}(t) - x_i(t) < v_i(t)$$

c) Randomization: $v_i(t) = \max(0, v_i(t)-1)$ with probability p .

d) Move: $x_i(t+1) = x_i(t) + v_i(t)$

where v_{\max} and p are the parameters of the traffic. The speed is decided based on rule a) primarily. Rules b) and c) are applied and the speed is overridden when the given conditions are met.

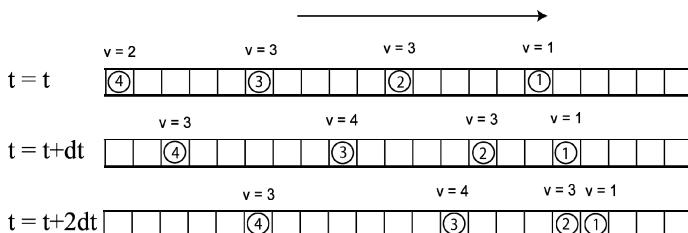


Figure 1. The pattern diagram of NS model.

B. Characteristics of Airport Traffic

NS model is a simple, easily-applied model, but it cannot express complicated airport traffic. Many factors should be considered to reflect the unique characteristics of airport traffic as listed below.

- 1) Take-off and landing
- 2) Aircraft separation
- 3) Aircraft dynamics
- 4) Algorithm for speed decision

5) Crossing

1) A take-off and a landing are the most important keys for airport traffic. At any time, a runway is usually used by a single aircraft only while other aircraft wait for take-off at the taxiway. Aircraft waiting to land are given priority to use the runway, which delays take-offs even further. To deal with congestion problems, most major airports have multiple runways, with each runway being used for either take-off or landing. Since congestion usually has a greater impact on take-off aircraft, only aircraft for take-off are considered in this paper.

The time separation between take-off aircraft is usually not constant, mainly due to wake turbulence. As every take-off aircraft induces wake turbulence, the following aircraft cannot take off until the wake turbulence becomes negligible. Bigger aircraft induce stronger wake turbulence; the time separation for take-off varies with the type of aircraft pair. Other factors play a role, too, e.g., the route after take-off, pilot judgment, aircraft speed. As for the route after take-off, wake turbulence should be considered for during flight, too, so if a pair of aircraft goes in the same direction after take-off, enough separation between the aircraft should be maintained. In terms of pilot judgment, wake turbulence cannot be seen, so the pilot tends to wait up to a minute or so on the runway when he feels the time separation is not enough. These factors should all be taken into account to estimate the precise time separation for take-off.

2) Aircraft separation is the distance between a pair of aircraft at a taxiway. As for car driving, the separation between two cars can be less than 1 meter, but this does not hold for aircraft. The separation between aircraft is in the hands of a pilot, and it varies with time.

3) Aircraft dynamics deals with the problem that aircraft have less agility than cars. The taxiing speed is usually controlled by engine thrust, which has a slow response. Although there is a braking system installed in order to decrease speed, this system is seldom used and aircraft usually decrease speed gradually by drag.

4) Algorithm for velocity decision considers the differences and similarities between aircraft and automobile in terms of velocity decision. Drivers usually rely on visual cues from the preceding car only to decide and control their car's velocity. On the other hand, aircraft can get information about the current status of congestion at the airport. If the pilot knows that the airport is congested, the aircraft will go taxiing slower than usual. Furthermore, thanks to the wide field of view of the pilot, they are aware of relative position of a lot of aircraft.

5) Crossing is not a distinct characteristic of an airport, but it should be considered. If the aircraft go through a crossing without turning, their speed is not reduced. However, when turning is necessary, the speed is reduced to a certain value. At the same time, the trajectories of turning and going straight are also different. In addition, as the order of take-off aircraft is decided by the control tower, an aircraft may have to wait before the crossing until the preceding aircraft passes the crossing.

In order to model the airport traffic, all of the above factors should be considered. However, the more complex the model,

the higher the number of parameters and lower the generality. The model should be constructed taking these factors into account, too.

III. PROPOSED AIRPORT TRAFFIC MODEL

In the previous section, NS model as explained, and five factors accounting for airport traffic were discussed. In this section, each factor is converted to several model rules, which are set in the model.

A. Take-off and Landing

In this paper, only take-off aircraft are considered. As mentioned before, several factors should be considered for the time separation between take-off aircraft. Here, a pair of aircraft and pilot judgment are considered. Firstly, a pair of aircraft is explained. The time separation for take-off depends on the size of aircraft, so heavy, medium, and light aircraft types are considered. Heavy aircraft are represented by B747, B777, A340; medium by B767, A300, and light by B737, A320. Take-off separation (t_{sep_1}) is defined by the following expression. t_{min} is the minimum time separation for take-off, and Δt is defined in Table I.

$$t_{sep_1} = t_{min} + \Delta t \quad (1)$$

TABLE I. CONDITIONS OF Δt

following/preceding	Heavy/Medium	Light
Heavy/Medium	t_0	0
Light	t_1	t_0

Note that heavy and medium aircraft fall in the same category in terms of Δt , so two parameters (t_0, t_1) are needed.

Next, the pilot judgment factor is explained. This factor indicates that the pilot waits for the wake turbulence effect to abate. Therefore, time separation (t_{sep_2}) is defined by the following expression.

$$t_{sep_2} = \begin{cases} d + \Delta d & (d < d_{limit}) \\ d & (d \geq d_{limit}) \end{cases} \quad (2)$$

where d is the time separation between the preceding aircraft take-off time and following aircraft arriving on the runway time, and Δd is the pilot judgment with respect to d_{limit} . Two parameters ($\Delta d, d_{limit}$) are necessary. Total time separation (t_{sep}) is defined as follows.

$$t_{sep} = \max(t_{sep_1}, t_{sep_2}) \quad (3)$$

The aircraft has to wait before entering the runway unless the take-off time separation is met. To describe this condition, the following rule is added on the rule b) in the NS model.

$$v_i(t) = 0 \quad \text{if the take-off timeseparation is not met} \quad (4)$$

B. Aircraft Separation

Aircraft separation sometimes varies throughout time, but in this paper, it is assumed to be constant. Therefore, the minimum distance separation between two aircraft is a constant number of cells (x_{min}).

C. Aircraft Dynamics

As previously mentioned, the braking system is rarely used during taxiing. However, according to NS model, deceleration is unlimited although acceleration is at most 1 cell per 1 time unit. Therefore, the expression of crash avoidance in NS model is revised, i.e., the maximum deceleration is set to 1 cell per 1 time unit.

Let me assume that the aircraft needs x units of separation and the current speed is v . The aircraft has to start decelerating in order to decelerate by 1 cell per 1 time unit if the following expression is satisfied.

$$x \leq \sum_{s=1}^v s = \frac{v(v+1)}{2} \quad (5)$$

Therefore, crash avoidance expression in NS model is replaced by the following expression. The original crash avoidance term is still kept to account for the case when an aircraft appears on the runway suddenly. This condition replaces rule b) in the NS model.

$$v_i(t) = \min(v_i(t-1) - 1, x_{i-1}(t) - x_i(t) - x_{min}) \quad (6)$$

$$\text{if } x_{i-1}(t) - x_i(t) - x_{min} \leq v_i(t-1)(v_i(t-1) + 1) / 2$$

D. Algorithm for Speed Decision

The aircraft speed is determined by the level of airport congestion. Needless to say, the separation from the preceding aircraft is important, but the aircraft speed is also affected by many other factors. Consequently, it is assumed that the aircraft speed is affected by all aircraft which are on the way to take-off. The closer the aircraft, the larger the impact on speed decision. In order to describe mathematically such an effect, the floor field model is introduced[8]. The floor field model was originally developed to express the long-range interaction, and in this example, the 2-dimensional dynamics of pedestrians is modeled using cellular automata. In the floor field model, each cell is assigned a value which keeps the cell status and memorizes the past information. This value is called floor field. The floor field model consists of the dynamic floor field (for past and long-range interaction information) and the static floor field (for cell status), here, only the dynamic floor field is considered. The dynamic floor field is modified by the

presence of pedestrians and has its own dynamics, i.e. diffusion and decay. Each pedestrian leaves a “trace”, so the floor field of the occupied cells increases. The floor field is decayed and delivered to the next cell, and thus the previous presence of pedestrian can be accounted for. In this paper, the dynamic floor field is applied, assuming that each aircraft leaves a trace. According to the floor field which are on the way to take-off, the aircraft speed is determined. The floor field and the aircraft speed are calculated based on the following expressions.

Trace: (x_i is the current i th aircraft position)

for $k = 0 : v_i(t-1)$

$$F(x_i - k, t) = F(x_i - k, t) + f_{trace} / (v_i(t-1) + 1) \quad (7)$$

Runway:

$$F(x_{runway}, t) = F(x_{runway}, t) + f_{runway} \quad (8)$$

Reference of aircraft speed :

$$v_i^{ref}(t) = \max \left(v_{\max} - \frac{\sum_{ontheway} F(x, t) \exp(-d/k)}{\sum_{ontheway} \exp(-d/k)}, 1 \right) \quad (9)$$

Acceleration and deceleration:

$$v_i(t) = \begin{cases} v_i(t-1) + 1 & \text{if } v_i^{ref}(t) \geq v_i(t-1) + 1 \\ v_i(t-1) & \text{otherwise} \\ v_i(t-1) - 1 & \text{if } v_i^{ref}(t) \leq v_i(t-1) - 1 \end{cases} \quad (10)$$

Decay:

$$F(x, t+1) = \delta F(x, t) \quad (11)$$

$F(x, t)$ indicates the floor field at the position x and time t . δ indicates the decay parameter. f_{trace} is a constant value which expresses “trace” effect. f_{runway} is also a constant value which expresses the “trace” effect at the runway because aircraft tend to decrease their speed as they approach the runway. It is assumed that the aircraft taxiing route is fixed in advance and the reference of aircraft speed ($v_i^{ref}(t+1)$) is calculated considering all floor fields which are on the way. According to the reference speed, the aircraft speed is determined. d indicates the number of the cells from the current position. The closer the floor field is, the bigger its effect is. k is the speed parameter, and a capital k indicates that the aircraft speed is affected strongly by the long-range floor field. Note that only the decay process is considered. Moreover, if the aircraft approaches the runway, the aircraft tends to decrease the speed, so the runway always leaves an extra of the floor field. These rules replace rule a) in the NS model.

E. Crossing

At a crossing point, the aircraft may need to decrease the speed, so the maximum speed at a crossing point v_{curve} is defined. In order to conduct a smooth deceleration, the deceleration should begin under the following condition.

$$d_{curve} \leq \sum_{s=v_{curve}+1}^v s = \frac{v(v+1)}{2} - \frac{v_{curve}(v_{curve}+1)}{2} \quad (12)$$

d_{curve} is the distance to the nearest curve from the current position. The following rule is added to the rule b) in the NS model.

(6)

$$\begin{aligned} v_i(t) &= v_i(t-1) - 1 \\ \text{if } d_{curve} &\leq v_i(t-1)(v_i(t-1)+1)/2 - v_{curve}(v_{curve}+1)/2 \end{aligned} \quad (13)$$

F. Others

According to actual data, additional aspects should be considered. Although the aircraft which is close to the runway decrease the speed according to the runway floor field term, the aircraft actually accelerate when no other aircraft is waiting to take off before the aircraft. Therefore, if the runway is not occupied and no aircraft is on the way to take-off, the speed is calculated with the notion that the floor field is equal to zero.

G. Summary of the rules

Several rules are applied in order to adapt the NS model to the airport surface simulation. This time, randomization is not applied. The rules are summarized as follows. As the NS model, rules b) and c) are applied and the speed is overridden when the given conditions are met.

a) Speed decision and trace calculation

Eqs. (7), (8), (9), (10), and (11).

b) Crash avoidance and smooth deceleration

Eqs. (6) and (13)

c) Take-off waiting

Eq. (4)

d) Move

$$x_i(t+1) = x_i(t) + v_i(t)$$

H. Parameter Tuning

In order to conduct a simulation, many parameters defined above have to be set. Actually, the parameters are t_{\min} , t_0 , t_1 , Δd , d_{limit} , x_{\min} , δ , f_{trace} , f_{runway} , k , v_{\max} , v_{curve} . All parameters are identified through simulations. Actual airport taxiing data is acquired first, and the taxiing start time/position and take-off time and the taxiing route is obtained for each aircraft. Through simulations, assuming that the initial time, the initial position, and the taxiing route are fixed, take-off time can be calculated. Therefore, the parameters which minimize the difference between actual take-off time and simulation should be chosen. The objective function is defined as follows:

$$J = \sqrt{\frac{1}{n} \sum_{i=1}^n (t_i^{act} - t_i^{sim})^2} \quad (14)$$

t_i^{act} and t_i^{sim} are the actual take-off time and that in simulation for i th aircraft, respectively. J has the unit of time, which can be used to evaluate the simulation accuracy. It is called a time index.

Finally, the factors discussed in this section are summarized in Table II.

TABLE II. THE FACTORS CHARACTERIZED FOR AIRPORT TRAFFIC IN THIS PAPER.

Factors	Details
Take-off and landing	The take-off time separation is determined by a pair of the aircraft size. The pilot instincts about wake turbulence are also considered.
Aircraft separation	The aircraft separation is a fixed number of cells, set by parameter tuning.
Aircraft dynamics	The maximum value of both the deceleration and the acceleration is set to 1 cell per 1 time unit.
Algorithm for speed decision	Using the floor field model, the aircraft speed is determined by the past and long-range interaction information which are on the way to the runway.
Crossing	The maximum speed at a crossing point is set.
Others	If the aircraft is too close to the runway, the aircraft speed follows a different rule.
Deceleration randomization	Not implemented.

IV. SIMULATION RESULTS

A. Airport and Data Acquisition

In order to confirm the validity of the proposed modeling method, airport surface data is necessary. In Japan, Tokyo (Haneda) International Airport has recently installed multilateration system, a surveillance system for the airport surface aircraft. The position of each aircraft can be obtained every second with an accuracy of 7.5 meters according to the specification of ICAO A-SMGCS manual[9]. Fig. 2 shows the map of Haneda airport. During certain hours, the runway is used for either take-off or landing, and this time, the runway (shown in blue) and the aircraft which take off from right to left, shown in the figure, are considered. The green parts are the taxiing way, and the yellow parts are the crossings where the maximum speed is equal to v_{curve} with the exception when aircraft go straight through the yellow part. Note that taxiways are divided in 5 meters cells, and the corners consist of two straight lines (not an arc).

The data used in this research was obtained between 5 pm and 9 pm on July 26th 2007. 114 aircraft took off during this period. The data is divided into two parts (A: 5:00 to 6:30 for 39 aircraft, B: 6:30 to 8:40 75 aircraft). The parameters are set by data B only, and the simulation result is discussed by both data A and B. Data B is treated as teaching data, and Data A as validation data. Note that the time interval in each simulation is 5 seconds. The maximum and average taxiing times among the aircraft are 1120 s and 502 s, respectively.

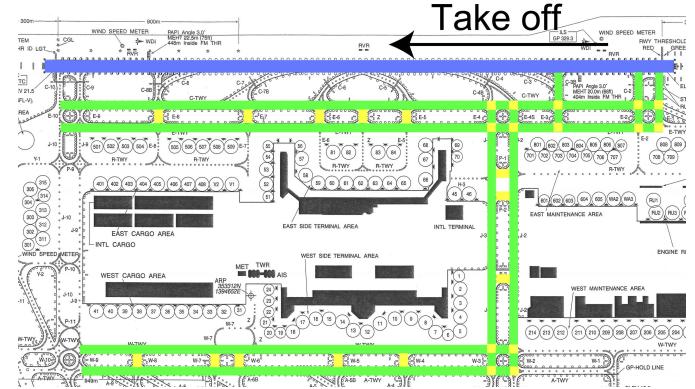


Figure 2. Haneda airport.

B. Parameter Settings and Simulation Results

The parameters are set to minimize the objective function. Firstly, the ranges of the parameters are determined, and in each combination of parameters, the objective function is calculated through a simulation. As the ranges of the parameters are adjusted by trial and error, the obtained parameters are not necessarily optimized. However, the ranges of the parameters are easily estimated to some extent, so it is considered that the parameters can be very close to the optimal ones.

Firstly, the optimized parameters are shown in Table 3. Using these parameters, a simulation is conducted, where the time index for data B is 28.96 s. The average taxiing time for data B is 502 s, and the normal take-off time separation is about between 80 and 120 s. The time accuracy of simulation is less than a half of take-off interval and 5 % of the taxiing time. Moreover, using the same parameters, the time index for validation data A is 27.87 s, where the parameters are not set based on this data. The accuracy is almost the same as that of data B, which indicates that this model is general enough..

TABLE III. THE OPTIMIZED PARAMETERS

t_{min}	70 s	Δd	30 s	δ	0.93	k	50 cells
t_0	15 s	d_{limit}	90 s	f_{trace}	35.5	v_{max}	12.75
t_1	40 s	x_{min}	25 cells	f_{runway}	4.0	v_{curve}	4.0

However, the time index is only the function of the difference between the real take-off and simulated take-off time. Therefore, the time histories of the aircraft trajectory are also shown. Fig. 3 shows the time histories of each aircraft trajectory when the airport is not so crowded. The horizontal axis indicates the time [second] from 0:00, and the vertical axis indicates the number of the cells to the runway. The black dots show the simulation result, and the red dots show the actual aircraft data. Note that each aircraft has a different initial position and follows a different route, so the trajectories sometimes cross each other. According to the figure, some aircraft in the simulation go faster than those in actual data especially when they are far from the runway. As for the rest,

most of the aircraft trajectories agree with the actual trajectory, which means the simulation works well to some extent.

Fig. 4 shows the time histories of each aircraft trajectory during a congestion. The figure format is the same as the previous one. During this period, since many aircraft go to the runway, its capacity is exceeded. This is when our model is most advantageous. While the first aircraft proceeds to the runway relatively smoothly, the following aircraft are gradually delayed. The 10th aircraft has to decrease its speed when it is as far as 200 cells from the runway, which verifies the simulation. In order to validate the speed decision algorithm, simulations not taking into account the floor field were conducted. These results are shown in Fig. 5. It is clearly seen that all aircraft move at the maximum speed until the minimum separation to the following aircraft is infringed. It should also be noted that the last aircraft took off about 200 s earlier than actual data nevertheless the take-off time separation is considered. This implies that the runway capacity depends on the airport congestion, too.

These results show that the proposed model has a great potential to model airport traffic even when the airport is congested. Since each aircraft has uncertainty during taxiing, the model cannot simulate all phenomena. However, by considering and implementing additional factors, the model can be further developed and improved. Moreover, the data used in the simulations presented in this paper is limited to a single day at a single airport and runway. Thus, extended simulations in various conditions are needed to confirm the general validity of this model.

V. CONCLUSIONS

So far, the airport surface congestion problem was not been examined in detail even though it can be critical for the future 4D trajectory concept. This paper focused on the airport congestion phenomenon, which was simulated based on the Nagel-Schreckenberg (NS) model. NS model was originally developed to describe highway car driving congestions, which differed significantly from aircraft ground congestions. In order to adopt this model to airport surface traffic, several key factors characterizing airport traffic were extracted. Then, they were implemented into the rules in the model. The simulation results indicated that the airport surface traffic was simulated well under the scenario of a congested airport. However, several improvements to the model still need to be made. Currently, the parameter decision process is based on trial and error, which is not effective enough. Besides, the rules characterizing the airport traffic are carefully chosen. Some current rules can be omitted and new rules should be introduced. It should also be verified how different the estimated parameters in the case of different runways, wind conditions, days, etc.

However, despite all these minor imperfections of the proposed model, I believe that it will contribute to the more exact time estimation of airport ground traffic and help the air traffic controllers manage the ground traffic more effectively.

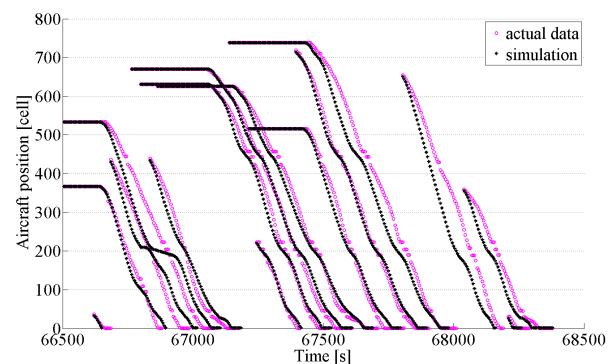


Figure 3. Simulation result for the not congested case.

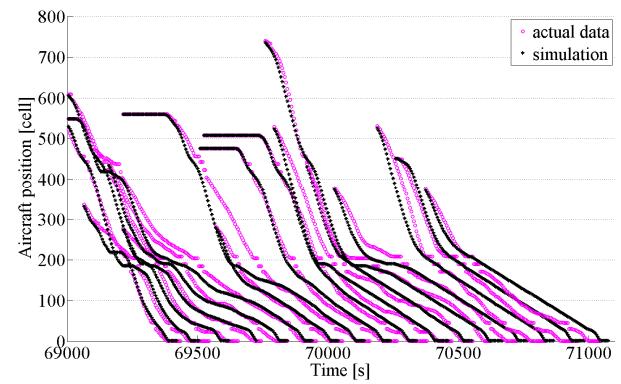


Figure 4. Simulation result for the congested case.

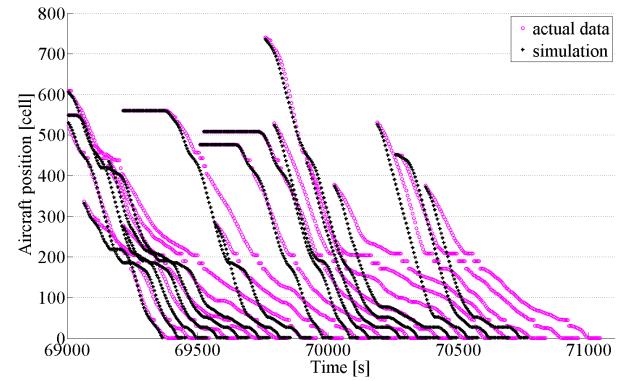


Figure 5. Simulation result for the congested case. (floor field is not considered)

ACKNOWLEDGEMENT

The author thanks for Japan Civil Aviation Bureau for providing the multilateration data at Tokyo (Haneda) International Airport.

REFERENCES

- [1] Gracia, J., Berlanga, A., Molina, J. M., and Casar, J. R., "Optimization of Airport Ground Operations Integrating Genetic and Dynamic Flow Management Algorithms," AI Communications, Vol. 18, No. 2, pp. 143-164, 2005.

- [2] Roling, P. C., and Visser, H. G., "Optimal Airport Surface Traffic Planning Using Mixed-Integer Linear Programming," *International Journal of Aerospace Engineering*, Vol. 2008, No. 1, 2008.
- [3] Atkins, S., Jung, Y., Brinton, C., Stell, L., Carniol, T., and Rogowski, S., "Surface Management System Field Trial Results," *AIAA 4th Aviation, Technology, Integration and Operations Forum*, AIAA-2004-6241, 2009.
- [4] Gong, C., "Kinematic Airport Surface Trajectory Model Development," *9th AIAA Aviation Technology, Integration and Operations Conference*, AIAA-2009-7076, 2009.
- [5] Bando, M., Hasebe, K., Nakayama, A., Shibata, A., and Sugiyama, Y., "Dynamical Model of Traffic Congestion and Numerical Simulation", *Physical Review E*, Vol. 51, pp. 1035-1042, 1995.
- [6] Nagel, K., and Schreckenberg, M., "A cellular automaton model for freeway traffic," *J. Phys. I France*, Vol. 2, No. 2, pp. 2221-2229, 1992.
- [7] Esser, J., and Schreckenberg, M., "Microscopic Simulation of Urban Traffic Based on Cellular Automata," *International Journal of Modern Physics C*, Vol. 8, No. 5, pp. 1025-1036, 1997.
- [8] Burstedde, C., Klauck, K., Schadschneider, A., and Zittartz, J., "Simulation of Pedestrian Dynamics Using a 2-dimensional Cellular Automaton," *Physica A: Statistical Mechanics and its Applications*, Vol. 295, No. 3-4, pp. 507-525, 2001.
- [9] International Civil Aviation Organization, *ICAO A-SMGCS manual*, §4.2.2.

Iterative Planning of Airport Ground Movements

Charles Lesire
 ONERA - DCSD
 Toulouse, France
 Email: charles.lesire@onera.fr

Abstract—Optimization of ground traffic is a major issue of air traffic management: optimal ground circulation could decrease flight delays and consequently decrease costs and increase passenger wellness. This paper proposes a planning algorithm for ground traffic based on contract reservation. This algorithm is iterative: it plans aircraft itinerary one after the other. A first version is described using the classical A^* algorithm. Then the model is extended to deal with time and speed uncertainty to ensure the feasibility of the planned trajectories while avoiding conflicts between aircrafts. Its efficiency is evaluated on Toulouse-Blagnac airport, regarding quality of the solution and computation times.

I. INTRODUCTION

One of the major issues of Air Traffic Management concerns the optimization of airport traffic. Indeed, the air traffic growth is having a hard impact on airport congestion. Flight delays are obviously impacted leading to an economic interest on ground traffic optimization methods. This optimization may also take into account ecologic issues such as noise and pollution reduction.

The optimization of ground traffic can hardly be performed by human controllers: managing several aircrafts moving on the airport during rush hours on quite complex taxiway networks may be difficult. It is especially the case when hard weather conditions occur (e.g. fog).

A lot of researches have tried to help ground controllers either by defining new visualization displays (DST [1], AMAN [2], DMAN, etc.) or by improving traffic predictability by sharing flight data between airports and controllers (CDM [3]). Currently, these methods help improving controller situation awareness or traffic predictability, but are not used to help planning the ground movements.

A lot of approaches manage flight departure scheduling from the airport using constraint relaxation [4], cooperative/coordinated plannings [5], [6], or optimization algorithms [7]. However, they do not consider prediction nor feasibility of the ground movements that correspond to these scheduling.

Some authors then tried to estimate taxiing time without planning or simulating the complete aircraft movements: [8] estimates this time using reinforcement learning; [9] stochastically computes flight delay based on airport congestion; [10] statistically estimates taxiing time from past data. These approaches could provide good approximations to schedule arrivals or plan air trajectories, but are not precise enough to estimate the pollution on the airport or control departure delays.

This paper presents an iterative algorithm for real-time planning of ground movements. This algorithm is intended to be used on-line to plan itineraries for aircrafts moving on an airport. Then these itineraries (sequel of points with time intervals) could be used either by human controllers, pilots, or by an automatic control law to control the aircraft speed along the trajectory. This algorithm is currently used in a simulation infrastructure allowing to evaluate airport capacities, environmental impacts, or optimization of new airport infrastructure.

Section II presents the overall problem and notations, and briefly describes the concepts. Section III details the A^* -based algorithm and some preliminary results. Then uncertainty management is addressed and experimented in section IV. Finally, section V discusses the benefits of the proposed approach, its limits, and the way it could be improved.

II. PROBLEM DESCRIPTION

A. Graph representation

The airport infrastructure is modelled as an oriented graph $G = (V, E)$ where vertices V are located points of the airport (taxiway intersections, gates, runway access points), and edges E are the airport taxiways. Each edge $(u, v) \in E$ has a weight corresponding to the length of the edge, i.e. $dist(u, v)$.

A flight f is described by a starting vertex v_s (a gate for departures, or a runway for arrivals), a final vertex v_f , a starting time t_s (the departure time from gate for departures, or the estimated landing time for arrivals), and a type or category, that will constrain the maximal speed s_{max} of the aircraft. Moreover, aircraft separation must be ensured: two aircrafts must never be closer than a given distance D .

B. Push-backs modelling

Departures usually follow a push-back procedure when leaving their gate. Such procedures are directly modelled in the graph structure by adding push-back nodes in the graph: the departure path from gates to push-back nodes are duplicated (Alg. 1, Fig. 1), allowing to define a reduced speed on push-back edges.

C. Problem and constraints

The problem is then to find, for each flight $f_k \in F$, an itinerary, or contract, i.e. a set of points and associated times $\sigma_k = (v_i, t_i)_{0 \leq i \leq l_k}$, such that:

Algorithm 1 *Duplicate(γ)*: Duplicate push-backs for departures from gate γ .

Require: $\gamma \in V$: an airport gate.

```

1: for all  $(u, v) \in E$ ,  $(u, v)$  push-back for gate  $\gamma$  do
2:   Create a copy  $u'$  of vertex  $u$ 
3:    $V \leftarrow V \cup \{u'\}$ 
4:    $E \leftarrow E - \{(\gamma, u), (u, v)\}$ 
5:    $E \leftarrow E \cup \{(\gamma, u'), (u', v)\}$ 
6: end for
```

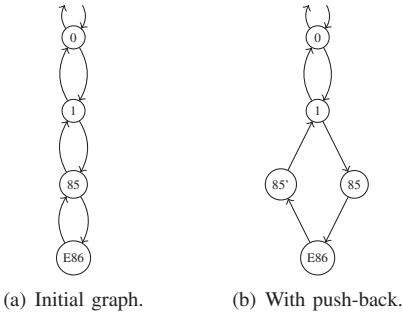


Fig. 1. Push-back nodes duplication: push-back of gate $E86$ is $(85, 1)$.

- first and last points correspond to the flight characteristics

$$v_0 = v_s, \quad t_0 = t_s, \quad v_{l_k} = v_f \quad (1)$$

- consecutive points are reachable

$$\forall i, (v_i, v_{i+1}) \in E \quad (2)$$

- the aircraft speed is below its maximal speed

$$\forall i, t_{i+1} > t_i \text{ and } s_k = \frac{\text{dist}(v_i, v_{i+1})}{t_{i+1} - t_i} \leq s_{max} \quad (3)$$

- aircraft separation is ensured

$$\begin{aligned} \forall f_j \in F, j \neq k, \forall v \in V, \forall t, t' / (v, t) \in \sigma_k, (v, t') \in \sigma_j, \\ |t' - t| \geq \frac{D}{s_k} \end{aligned} \quad (4)$$

The overall objective is to minimize the travel time of all the aircrafts:

$$\min \sum_{f_k \in F} t_{l_k} \quad (5)$$

Computing a solution to this problem is quite complex. Although finding a path for a given aircraft f in the airport graph could be efficiently done in $\mathcal{O}(|V|^2)$ – Dijkstra algorithm complexity – computing a global optimum while managing time constraints (including separation) worsen the complexity to $\mathcal{O}(|F|!|V|^4)$. This is merely intractable without any appropriate resolution method.

Gotteland [11] proposes a time-bounded approach in which the optimization process considers all flights during an horizon H_p . His approach optimizes the order in which flights must be planned, and their itineraries, to minimize the global delay. By considering all the flights, this approach is

still complex, and the author has to consider a limited search graph, leading to sub-optimal results. In [12], an iterative approach is proposed, but its complexity avoid to use it on real-time, or leads to the same sub-optimal considerations as [11].

The approach proposed in this paper decomposes the algorithm into iterative computations: each flight is planned one after the other. The contract of flight f_k is computed using the contracts of already planned flights without allowing to modify them. This solution is obviously not optimal regarding the global objective of equation (5). However, it is more realistic, as aircrafts start moving on the airport one after the other depending of their departure time. This approach is also robust to delays, as a flight starting δt after its initial starting time will not influence already planned flights but will try to be inserted in the current circulation.

III. ITERATIVE PLANNING ALGORITHM

A. A^* -based modeling and planning

As discussed before, the approach proposed in this paper is iterative. Each flight will be announced and planned one after the other depending on its starting time. The flight itinerary is planned according to already reserved contracts in order to satisfy the separation constraint.

The algorithm is based on A^* [13] (Alg. 2). It computes an itinerary from an initial node v_0 to a final node v_f . A^* is a best-first search algorithm, exploring nodes minimizing function $g + h$ where g is the cost function and h the heuristic. If h is admissible (it must not overestimate the real cost to the goal), A^* returns a solution minimizing g . Classically, h is the euclidean distance, or other norms (1-norm, ∞ -norm, ...) The optimal path is finally extracted reading the parent relation p from goal v_f back to the initial vertex v_0 .

Constraints (3) and (4) are not managed by the algorithm itself but by defining an appropriate cost function. In standard shortest-path problems, g is defined as the weight matrix of graph G , and $COST$ function is given by equation (6).

$$\forall (u, v) \in E, COST(u, v) = g(u) + dist(u, v) \quad (6)$$

In the ground movements problem, the aim is to minimize the travel time of each flight. Hence, the cost function of a node v_{i+1} must be expressed according to the time taken by the aircraft to move from the previous point v_i to v_{i+1} . Then $COST(v_i, v_{i+1}) = t_{i+1}$. Constraint (3) leads to:

$$\frac{\text{dist}(v_i, v_{i+1})}{t_{i+1} - t_i} \leq s_{max} \Leftrightarrow t_{i+1} \geq t_i + \frac{\text{dist}(v_i, v_{i+1})}{s_{max}} \quad (7)$$

providing a lower bound for t_{i+1} .

Constraint (4) is satisfied by Alg. 3. This algorithm computes the shortest time t_v at which the aircraft will be able to arrive at v while satisfying the separation constraint. $contract(f_j)$ is the contract already planned for flight f_j ,

Algorithm 2 The A^* algorithm.

```

1:  $\mathcal{O} \leftarrow \{v_0\}$ 
2:  $\forall v \in V, g(v) \leftarrow +\infty,$ 
3:  $g(v_0) = 0, h(v_0) \leftarrow h(v_0, v_f)$ 
4:  $\forall v \in V, p(v) \leftarrow v$ 
5: while  $\mathcal{O} \neq \emptyset$  do
6:    $x \leftarrow \operatorname{argmax}_{z \in \operatorname{argmin}_{y \in \mathcal{O}} (g(y) + h(y))} g(z)$ 
7:   if  $x = v_f$  then
8:     return shortest path from  $v_0$  to  $v_f$ 
9:   end if
10:   $\mathcal{O} \leftarrow \mathcal{O} - \{x\}$ 
11:  for all  $(x, y) \in E$  do
12:     $g'(y) \leftarrow COST(x, y)$ 
13:    if  $g'(y) < g(y)$  then
14:       $g(y) \leftarrow g'(y)$ 
15:       $p(y) \leftarrow x$ 
16:       $\mathcal{O} \leftarrow \mathcal{O} \cup \{y\}$ 
17:    end if
18:  end for
19: end while

```

Algorithm 3 Cost function $COST(u, v)$.

```

1:  $t_v = t_u + \frac{dist(u, v)}{s_{max}}$ 
2: for all  $f_j \in F, j < k$  do
3:    $t' = contract(f_j, v)$ 
4:    $\delta = \frac{D}{s_{uv}} = \frac{D}{dist(u, v)}(t_v - t_u)$ 
5:   if  $|t_v - t'| < \delta$  then
6:      $t_v = t' + \delta$ 
7:   end if
8: end for
9: return  $t_v$ 

```

giving for each node v a time t' at which the aircraft will pass over v .

Algorithm 3 is executed at each step of the A^* algorithm. Hence the complexity of the itinerary computation for a flight is $\mathcal{O}(|V|^2 |F|)$, where $\mathcal{O}(|V|^2)$ is the complexity of Alg. 2 and $\mathcal{O}(|F|)$ the complexity of Alg. 3.

The heuristic function is given by equation (8). This heuristic is admissible ensuring the optimality of Alg. 2.

$$h(v_i) = h(v_i, v_f) = \frac{dist(v_i, v_f)}{s_{max}} \quad (8)$$

B. Results

The previous algorithms have been implemented in C++, using the Boost Graph Library structures and algorithms. Some experiments have been made based on the Toulouse-Blagnac airport, whose graph has 205 nodes and 361 edges (Fig. 11).

Figure 2 shows the number of delayed flights (in %) according to the number of flights planned on the airport during 100 hours¹. Each flight start and final point is uniformly

TABLE I
RESULTS ON A ONE-DAY TRAFFIC PLANNING.

Flights per hour	10
Delayed flights (%)	10.39
Flights w. delay > 5%	1.3
Flights w. delay > 10%	0.65
Flights w. delay > 20%	0
Average delay (in %)	4.28
Worst delay (in %)	10.05

drawn from the set of gates or runways of the airport graph. The flight starting time is also uniformly drawn according to the number of flights managed during the 100h.

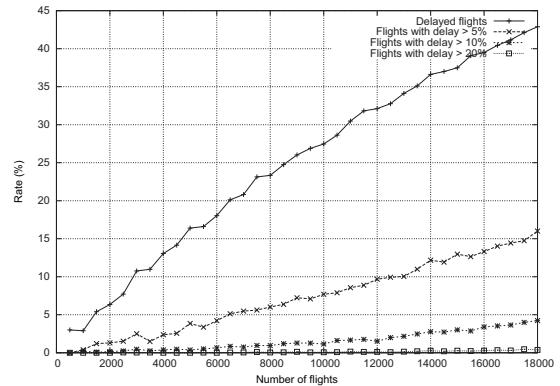


Fig. 2. Relative number of delayed flights.

The relative number of delayed flights (in % of the total number of flights) is linear, showing the complexity to manage a high number of aircrafts in such an airport. Results from an actual one-day traffic on Blagnac airport are shown in Tab. I.

The number of delayed flights is not consistent between random simulation results and the real traffic data. This can be explained by the fact that the real traffic is not uniform over the day. Rush hours are nearer to 25 fl/h (2500 flights in 100 hours), giving more consistent results (around 10% of flights are delayed).

Figure 3 shows the resulting average and maximal delays for delayed flights according to the number of flights. The average and worst delays are consistent with those of the real Blagnac traffic results.

Globally, the results for the Blagnac airport give some acceptable delays. Managing around 20 flights per hour leads to 8% delayed flights, with an average delay less than 5% of their travel time.

Moreover, the computation time associated to the itinerary planning is less than 1 second per flight on a Core2 2.16GHz, 2Go RAM standard laptop, which makes the process fully usable on-line.

However, the resulting itineraries, that correspond to sequels

¹This simulation time has been chosen to have statistically sound results.

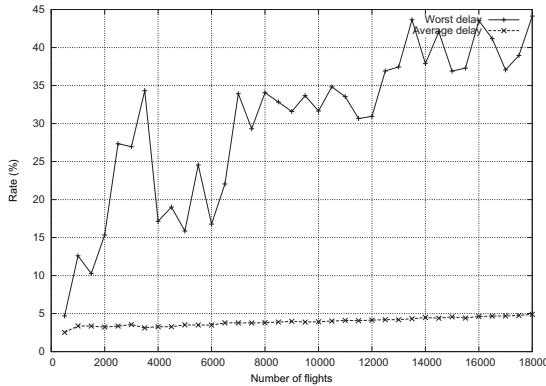


Fig. 3. Relative average and worst delays.

of timed nodes, are not realistic. The hypothesis is that the aircraft speed is constant on each edge, leading to a discontinuous speed evolution of the aircraft (Fig. 4) along its trajectory (Fig. 5).

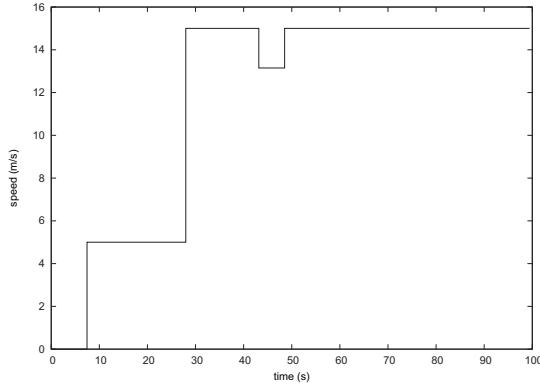


Fig. 4. Speed profile of flight 988.

The second drawback concerns the accuracy of starting time. To be sure an itinerary will be ready for an arriving flight as soon as it goes out of its runway, the planning process must compute its itinerary around a couple of seconds before it lands. However, the "starting time" (i.e. the time at which the aircraft will join the first taxiway) cannot be known precisely.

The following section deals with these two drawbacks and the way their associated uncertainties are managed in the planning algorithm.

IV. MANAGING UNCERTAINTY

Improving the realism of the planned itineraries means that the strong time constraint (a unique date associated to a node) must be relaxed. The itinerary must be represented as a sequel of nodes associated to time intervals. These intervals may be due to: (1) the uncertainty on the flight starting time (that will

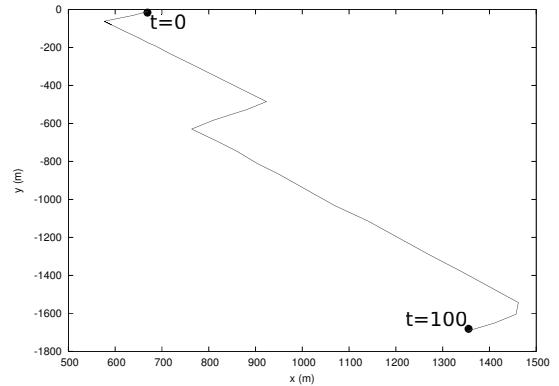


Fig. 5. Trajectory of flight 988.

TABLE II
ALLEN'S ALGEBRA RELATIONS.

Timeline	Relation	Notation ³
$\underline{\underline{X}}$	X before Y	$X < Y$
$\underline{\underline{Y}}$	Y after X	$Y > X$
$\underline{\underline{X}} \quad \underline{\underline{Y}}$	X meets Y	XmY
$\underline{\underline{Y}} \quad \underline{\underline{X}}$	Y is met by X	$YmiX$
$\underline{\underline{X}} \quad \underline{\underline{Y}}$	X overlaps Y	XoY
$\underline{\underline{Y}} \quad \underline{\underline{X}}$	Y is overlapped by X	YoI
$\underline{X} \quad \underline{Y}$	X starts Y	XsY
$\underline{Y} \quad \underline{X}$	Y is started by X	$YsiX$
$\underline{\underline{X}} \quad \underline{\underline{Y}}$	X finishes Y	XfY
$\underline{\underline{Y}} \quad \underline{\underline{X}}$	Y is finished by X	$YfiX$
$\underline{\underline{X}} \quad \underline{Y}$	X during Y	XdY
$\underline{\underline{X}} \quad \underline{Y}$	Y contains X	$YdiX$
$\underline{\underline{X}} \quad \underline{\underline{Y}}$	X equals Y	$X = Y$

be propagated over the itinerary), or (2) the uncertainty on the aircraft speed, leading to an uncertainty on the time taken to cover a taxiway (that will increase over the itinerary).

A. Propagating time uncertainty

To represent time uncertainty, the itinerary of flight f_k is now a set $\sigma_k = (v_i, T_i)_{0 \leq i \leq l_k}$, where T_i is an interval $[t_i^-, t_i^+]$. The cost function for the A^* algorithm must be defined to provide, for each node v_{i+1} , a time interval T_{i+1} during which² the aircraft can go over node v_{i+1} while satisfying separation constraint (4).

As done in Alg. 3, T_{i+1} is iteratively computed by comparing the sooner possible interval T to already planned contracts T' . This comparison is based on the Allen's algebra [14]. Allen defines thirteen relations to compare two intervals, summarized in Tab. II.

The fact that X is either (for instance) before or overlaps Y is noted $X\{<, o\}Y$.

Interval time computation is ensured by Alg. 4:

- The computation of the separation time is over-estimated to guarantee the separation constraint (line 4);

²Actually f_k can be on v_{i+1} at any time $t \in T_{i+1}$.

- If T_v does not intersect $T' + \Delta$, separation is ensured and T_v is not modified (line 6);
- If T_v has an intersection with $T' + \Delta$, and finishes later (line 8), then T_v is truncated: as the aircraft may arrive on v at any time between t_v^- and t_v^+ , it can obviously move slower to arrive between $(t_v^+ + \delta_T)$ and t_v^+ ;
- Line 10 is an extreme case of the previous one.

Algorithm 4 Interval cost function $COST(u, v)$.

```

1:  $T_v \leftarrow T_u + \frac{dist(u,v)}{s_{max}}$ 
2: for all  $f_j \in F, j < k$  do
3:    $T' \leftarrow contract(f_j, v)$ 
4:    $\delta_T \leftarrow \frac{D}{s_{min}} = \frac{D}{dist(u,v)}(\max(t_v^+, t'^+) - t_u^-)$ 
5:    $\Delta \leftarrow [-\delta_T, +\delta_T]$ 
6:   if  $T \{<, >\} T'$  then
7:     print
8:   else if  $T_v \{si, oi, di\} T' + \Delta$  then
9:      $T_v \leftarrow [t'^+ + \delta_T, t_v^+]$ 
10:  else if  $T_v \{s, f, fi, o, d, =\} T' + \Delta$  then
11:     $T_v \leftarrow [t'^+ + \delta_T, t'^+ + \delta_T]$ 
12:  end if
13: end for
14: return  $T_v$ 
```

In the special case where $T_u = [t_u, t_u]$ (i.e., is reduced to a single time) Alg. 4 is similar to Alg. 3.

B. Speed uncertainty

Managing starting time uncertainty gives some flexibility to the flight trajectories: arriving at a given node v must be done between t_v^- and t_v^+ , allowing the aircraft to manage its speed. However, it is not sufficient: T_v intervals may be reduced to singletons (Alg. 4, line 11), leading to a discontinuous speed profile.

Hence a speed uncertainty must be introduced in the $COST$ function to have a more realistic speed profile. This uncertainty is given by a δ_S parameter representing the tolerance over the nominal speed s_k . Typically, $\delta_S = 3\text{m/s}$ in the following experiments.

Algorithm 5 is a modified version of Alg. 4 that introduces speed uncertainty. Indeed, Alg. 5 manages both start time uncertainty and speed uncertainty, and the way this uncertainty is propagated (and evolves) along the itinerary.

The overall complexity has not changed ($\mathcal{O}(|V|^2 |F|)$), but the computation time should be slightly higher as interval operations are more expensive than float operations.

C. Results

Figure 6 shows the speed profile bounds (min and max speeds) for Flight 988 (see Fig. 5 for flight trajectory and Fig. 4 for its previous speed profile). While there still is a discontinuity around $y = 40$, the provided profile allows the aircraft speed to be more smoothly controlled. The itinerary is now more realistic and executable.

³i stands for inverse.

Algorithm 5 Interval cost function $COST(u, v)$ with speed uncertainty.

```

1:  $T_v \leftarrow T_u + \frac{dist(u,v)}{[s_{max}-\delta_S, s_{max}+\delta_S]} = T_u + [\frac{dist(u,v)}{s_{max}+\delta_S}, \frac{dist(u,v)}{s_{max}-\delta_S}]$ 
2: for all  $f_j \in F, j < k$  do
3:    $T' \leftarrow contract(f_j, v)$ 
4:    $\delta_T \leftarrow \frac{D}{s_{min}} = \frac{D}{dist(u,v)}(\max(t_v^+, t'^+) - t_u^-)$ 
5:    $\Delta \leftarrow [-\delta_T, +\delta_T]$ 
6:   if  $T \{<, >\} T'$  then
7:     print
8:   else if  $T_v \{si, oi, di\} T' + \Delta$  then
9:      $T_v \leftarrow [t'^+ + \delta_T, t_v^+]$ 
10:  else if  $T_v \{s, f, fi, o, d, =\} T' + \Delta$  then
11:     $T_v \leftarrow t'^+ + \delta_T$ 
12:  end if
13: end for
14: return  $T_v$ 
```

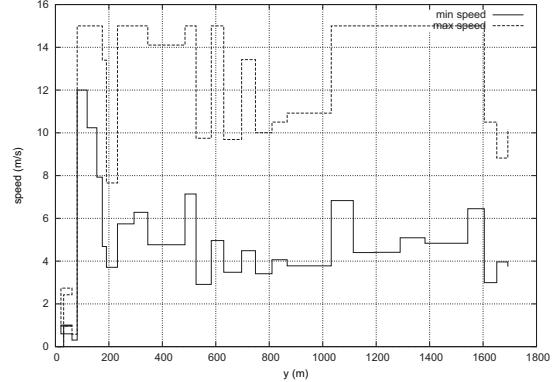


Fig. 6. Speed bounds for flight 988 along its trajectory.

Figures 7 and 8 present the evolution of the number of delayed flights and their delays according to the width of the starting time interval $|T_0|$. The number of delayed flights is near constant (Fig. 7), meaning that $|T_0|$ has only a local effect on "already delayed" flights. Moreover, although the maximal delay is linear according to $|T_0|$ – which is reasonable – the average delay is always under 20% (Fig. 8).

Figures 9 and 10 clearly show that speed uncertainty as very few influence on the number of delayed flights and their delays.

Table III shows results on the Blagnac airport actual traffic using a time interval uncertainty of 20 seconds and a speed uncertainty of 3 m/s. These results are encouraging regarding the number of delayed flights and their average delay. However, the worst delay, that correspond to an actual travel time more than fifth the optimal travel time, clearly emphasizes the major drawback of the proposed approach: itineraries are computed to satisfy aircraft separation whatever the other aircrafts trajectories, i.e. considering their worst possible delay.

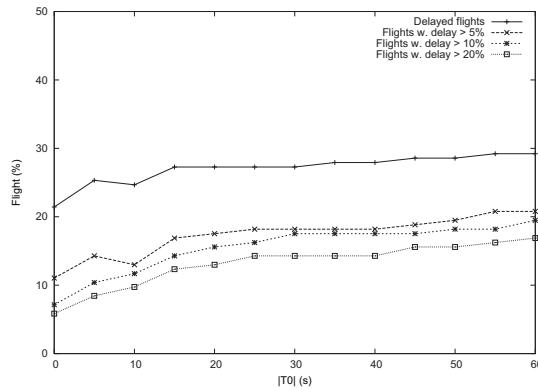


Fig. 7. Number of delayed flights according to the initial time uncertainty.

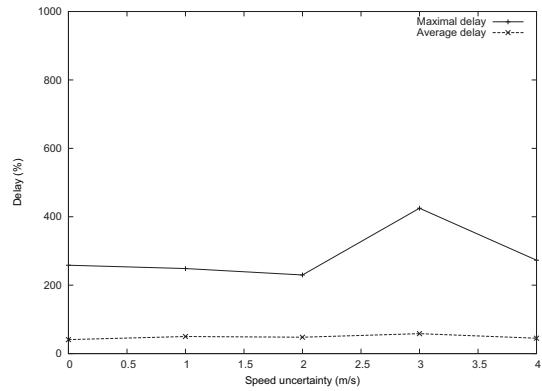


Fig. 10. Average and maximal delays according to speed uncertainty.

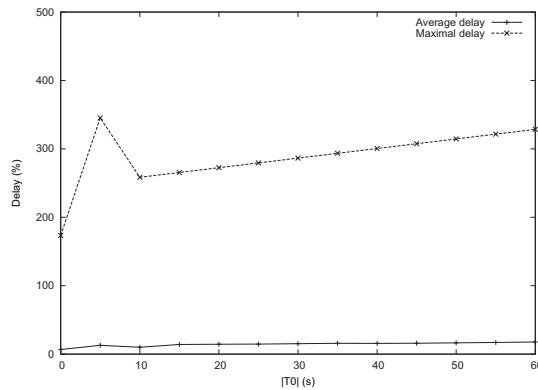


Fig. 8. Average and maximal delays according to the initial time uncertainty.

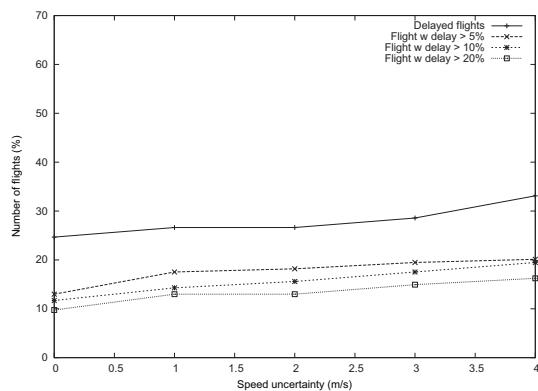


Fig. 9. Number of delayed flights according to speed uncertainty.

V. CONCLUSION

The approach proposed in this paper is dedicated to compute airport ground movements. The planning algorithm

TABLE III
RESULTS WITH $|T_0| = 20$ AND $\delta_S = 3$.

	Flights per hour	10
Delayed flights (%)	29.2	
Flights w. delay > 5%	19.5	
Flights w. delay > 10%	18.2	
Flights w. delay > 20%	14.9	
Average delay (in %)	17.8	
Worst delay (in %)	447.4	

is iterative, i.e. it plans flights one after the other, ensuring speed and separation constraints. Several cost function of A^* have been implemented to manage time and speed uncertainties as time intervals. The results have shown the realism of the provided itineraries (in term of delays, speed profile and airport capacity), and proved the efficiency of the algorithm in term of computation time (less than 1 second per flight).

However, some drawbacks must be pointed out:

- 1) Controlling the aircraft speed to ensure separation may lead to unexpected situations where the aircraft speed is very small; as separation constraint is only verified on nodes (and not on edges), a situation where several aircrafts are slowly moving on a busy taxiway is possible.
- 2) The planned trajectory are over-constrained: during execution, the aircraft will have a specific trajectory, arriving on each node at a unique time; next flights will not reconsider their itinerary and will then use a "worst-time" assumption.

These two issues will be addressed by adopting a real-time behaviour, each flight planning (and modifying) its itinerary while moving on the airport. Moreover such an approach may allow to deal with runway crossing (which is dependent on the actual situation and is not addressed in this paper), and on-line control clearances. These developments will

then include a simulation of the aircraft trajectory intimately connected to the planning algorithm.

Finally, the proposed approach is to be used not only to plan and simulate ground movements, but also to evaluate airports capacities, or give accurate estimation of "gate to runway" travel time to the departure management team or runway control.

ACKNOWLEDGEMENTS

This work is part of the IESTA program, funded by a set of European (ERDF - European Regional Development Fund) and national French public credits. The activities described in this paper are included in a collective work carried out by the whole team of the IESTA program. Moreover, the IESTA program owes its existence to a federative and close collaboration between several Onera scientific departments that gather a multi-disciplinary team of scientific experts of the wide range following domains: Long-Term Design & Systems Integration, Systems Control and Flight Dynamics, Computational Dynamics & Aeroacoustics, Physics, Instrumentation & Sensing, Aerodynamics & Energetics Modelling.

REFERENCES

- [1] S. Swierstra and S. Green, "Common trajectory prediction capability for decision support tools," in *ATM R&D Seminar*, Budapest, Hungary, 2003.
- [2] H. Oberheid and D. Soffker, "Designing for cooperation mechanisms and procedures for air-ground integrated arrival management," in *IEEE Conf. on Systems, Man, and Cybernetics*, Montréal, Canada, 2007.
- [3] P. Martin, O. Delain, and F. Fakhoury, "Collaborative decision making: results of experiments to identify limitations of information exchanges in stand and gate operations," in *ATM R&D Seminar*, Santa Fe, NM, USA, 2001.
- [4] P. van Leeuwen and B. van Hanxleden, "Scheduling aircraft using constraint relaxation," in *UK Planning and Scheduling Meeting*, Glasgow, UK, 2003.
- [5] H. de Jonge, E. Tuinstra, and R. Seljée, "Outbound punctuality sequencing by collaborative planning," NLR - National Aerospace Laboratory, The Netherlands, Tech. Rep., 2005.
- [6] D. Bohme, R. Brucherseifer, and L. Christoffels, "Coordinated arrival departure management," in *ATM R&D Seminar*, Barcelona, Spain, 2007.
- [7] R. Deau, J. Gotteland, and N. Durand, "Runways sequences and ground traffic optimisation," in *Int. Conf. on Research in Air Transportation (ICRAT'08)*, Fairfax, VA, USA, 2008.
- [8] P. Balakrishna, R. Ganesan, and L. Sherry, "Application of reinforcement learning algorithms for predicting taxi-out times," in *ATM R&D Seminars*, Napa, CA, USA, 2009.
- [9] R. Hoffman, M. Ball, R. Smith, and A. Mukherjee, "Ration-by-distance with equity guarantees: a new approach to ground delay program planning and control," in *ATM R&D Seminar*, Barcelona, Spain, 2007.
- [10] P. Pina and J. De Pablo, "Benefits obtained from the estimation and distribution of realistic taxi times," in *ATM R&D Seminar*, Baltimore, MD, USA, 2005.
- [11] J. Gotteland, N. Durand, and J. Alliot, "Genetic algorithms applied to airport ground traffic optimization," in *Congress of Evolutionary Computing*, Canberra, Australia, 2003.
- [12] C. Lesire, "Automatic planning of ground traffic," in *AIAA Aerospace Sciences Meeting*, Orlando, FL, USA, 2009.
- [13] P. Hart, N. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," *IEEE Transactions on Systems Science and Cybernetics*, vol. 4, no. 2, 1968.
- [14] J. Allen, "Maintaining knowledge about temporal intervals," *Communications of the ACM*, vol. 26, no. 11, 1983.

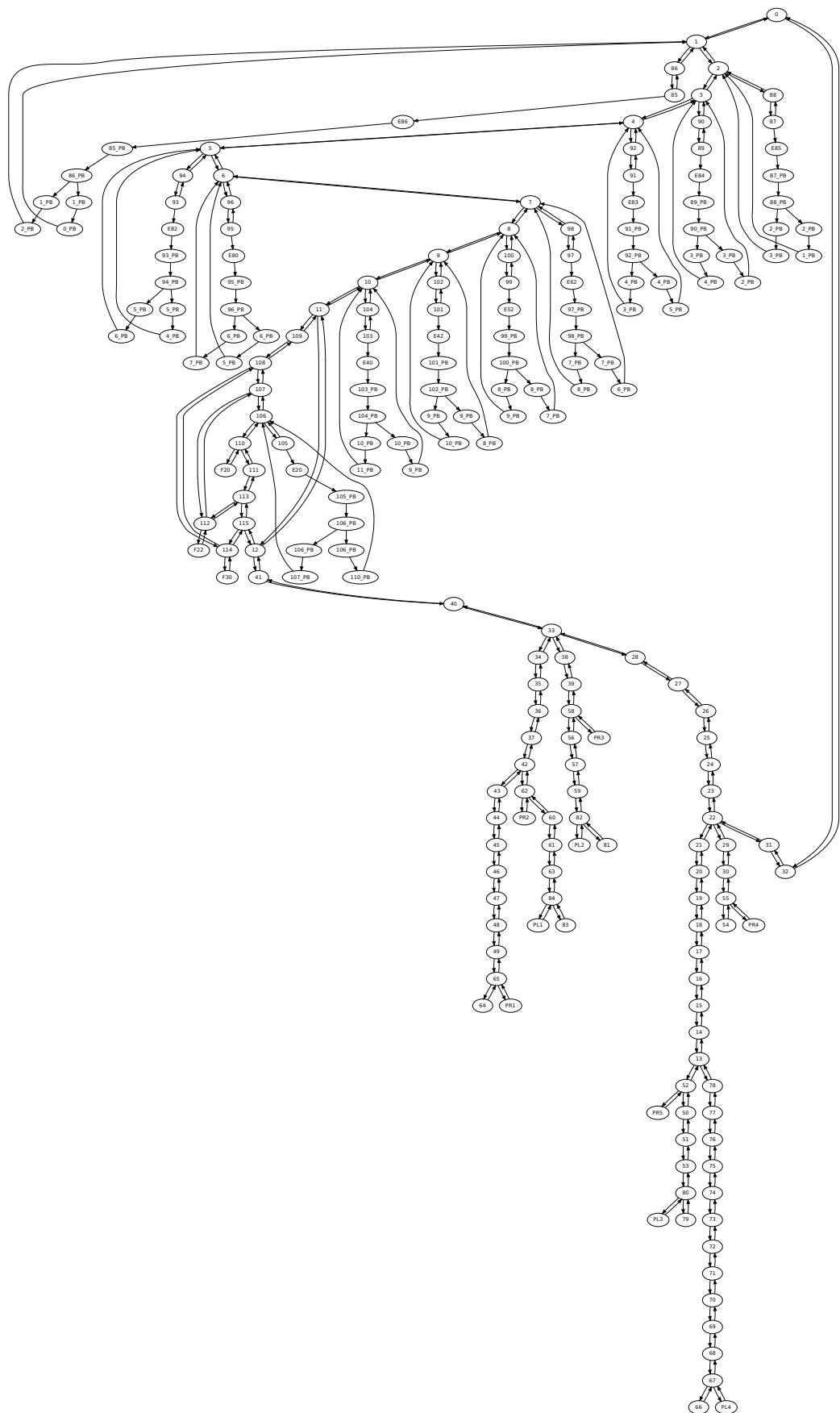


Fig. 11. The Toulouse-Blagnac airport graph.

Potential of Dynamic Aircraft to Runway Allocation for Parallel Runways

Martin Fritzsche, Thomas Günther, and Hartmut Fricke

Chair of Air Transport Technology and Logistic

Technische Universität Dresden, Germany

fritzschemartin@aol.com; guenther@ifl.tu-dresden.de; fricke@ifl.tu-dresden.de

Abstract – A flexible and demand-driven utilisation of available runway infrastructure plays an important role to meet aviation's future targets regarding capacity, efficiency and environmental sustainability. This paper presents and validates a heuristic algorithm to dynamically allocate arrival aircraft to one of two parallel runways. It considers both ATC regulations and modelled preferences of the airspace user and airport operator. It is designed to balance runway loads to reduce arrival and departure delays, taxi times, resulting fuel consumption and aircraft emissions. Particular focus is also set on ATC controller workload to avoid negative effects on safety. The implemented algorithm was applied in a set of fast-time simulations for the new Berlin Brandenburg International Airport (BBI). It promises significant operational potential, especially but not exclusively for airports with independent parallel runways.

Keywords: Runway Allocation, Arrival Management, Airport Capacity, Efficiency, Environmental Sustainability.

I. INTRODUCTION

The modernisation of the Air Traffic Management (ATM) system in Europe within the frame of the Single European Sky ATM Research Programme (SESAR) requires an intensified partnership and calls for a collaborative decision making (CDM) between all involved partners. This is a major premise in order to cope with the performance targets for the year 2020, forecasting a 3-fold increase in air traffic demand and promising the reduction of aircraft emissions by 10 percent [1]. It is expected that airports will remain a crucial capacity element in the ATM system. As such, the efficient utilisation of the existing runway infrastructure is a dominant asset. This can be equally as effective as expanding airport infrastructure, without incurring negative financial, societal and environmental costs.

Runways are a vital component within the ATM system for enabling a user-orientated flight trajectory planning and execution [2]. The SESAR Target Concept [3] addresses this requirement in regard to runway management during the execution phase. Improvements in runway throughput, utilisation and safety shall be achieved by implementing operating procedures that balance actual demand and capacity, minimize traffic queues, de-conflict and separate traffic and apply safety nets. The future ATM-system shall combine strategic traffic flow management with tactical air traffic control also within the Terminal Manoeuvring Area (TMA).

Although a fixed arrival and departure route structure within the TMA will remain inevitable due to even higher traffic complexity and environmental constraints, the route design shall nevertheless allow for dynamic adaptations according to the actual traffic situation. In that way an efficient and individually adjustable runway allocation scheme can contribute to enabling allocation of traffic demand to the available airport capacity as flexible as possible.

The current paper complements the manifold research performed in the context of arrival and departure traffic optimisation. Currently, focus of according support systems (i.e. AMAN, DMAN) is primarily set to the allocation of target times with regard to flow control measures at airports and the surrounding airspace. The results of several studies (e.g. [4]; [5]) reveal the potential for efficiency benefits. In addition, this paper identifies the potential for a decision support tool that provides specific runway allocation suggestions for arrival traffic. The presented algorithm combines various criteria for each particular flight regarding the current traffic and capacity situation. It is designed to fulfil the above mentioned general objectives for efficient runway management.

II. BACKGROUND

This study is motivated by the planning process for an optimised runway concept for Berlin Brandenburg International Airport (BBI) which is currently under construction. The concept for dynamic runway allocation derives from operational ATC-procedures that are already applied at airports with parallel runways, operated in mixed-mode (i.e. Munich Airport) [6], [7]. However, the concept can also be adapted to the operational constraints of airports with dependent parallel runways (i.e. Frankfurt/Main Airport).

A. Operational Procedures

The most prevailing factor for runway allocation of arrival traffic is the geographical origin from where the aircraft enters the TMA. At Munich TMA¹ all aircraft entering via the northern metering fix points follow pre-defined standard

¹ All described procedures of Munich Approach Control derive from personal observation, explanations by ATC-controllers and management of the DFS (German Air Navigation Service Provider) as well as national ATC-regulations.

arrival routes (STAR), if not vectored manually by the Pickup Controller. When reaching the downwind leg, pilots are advised by the Feeder Controller to turn base and intercept the precision (ILS) segment of the northern runway (see Fig. 1). The same procedure is applied accordingly for traffic entering via the southern metering fix points and landing on the southern runway [8], [9].

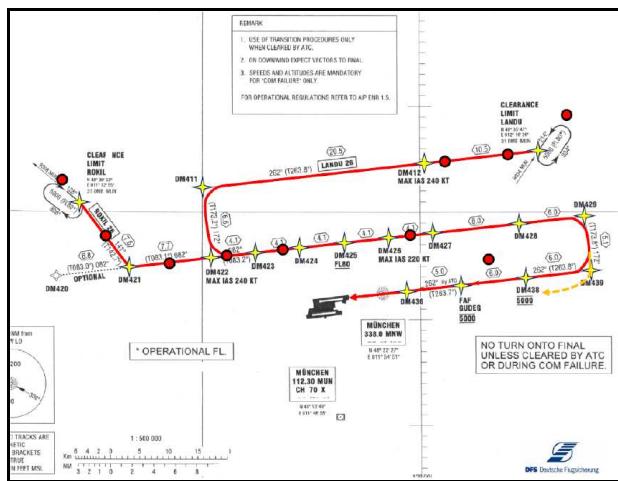


Figure 1. Arrival-Transition for Runway 26R, Munich Airport [9]

In this way conflicting traffic situations on base and on final are vastly resolved. However, the operational practice reveals exceptions to this rule. It can be observed that a minority of flights are being advised by the responsible Feeder Controller to land on an alternate runway. In this case aircraft fly an extended base leg and turn on final of the parallel runway (see dotted line in Fig.1). The analysis of radar flight track data, as shown in Fig. 2, covering an exemplary period of two hours of arrival traffic at Munich Airport, shows that a significant amount of traffic is being allocated to a different runway with the majority of flights still following the standardised approach. At Munich Approach Control one Feeder Controller is responsible for the airspace that covers the downwind, base and final legs of both runways, as illustrated in Fig. 2. He is responsible for separating traffic and establishing an efficient final approach sequence.

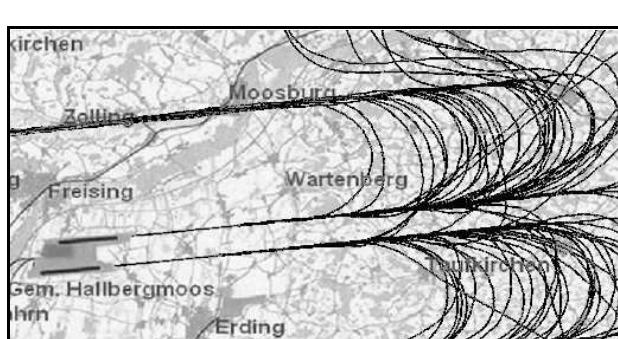


Figure 2. STANLY track data (2h) for arrival traffic at Munich Airport

Most re-allocations are initiated on request by pilots upon entering the TMA in order to reduce taxi time to the expected parking position near a given runway. Moreover, the Feeder Controller makes tactical sequencing decisions, leading to re-allocations, in order to balance the arrival flow for both runways or to create gaps in the arrival sequence for departure traffic on one of the runways. Since this is the responsibility of only one controller, no additional communication workload between ATC units is required. This may change once the responsibility would be shared between two Feeder Controllers for the northern and southern runway.

B. Motivation

The allocation of traffic according to the geographical direction upon entering the TMA complies with the necessity of predefined standard operating procedures for safety reasons, in particular under adverse weather conditions, reduced capacity or emergency situations. However, operating concepts, as described above, provide certain flexibility in the operational process to allocate individual aircraft to routes and runways approximately 30min or 15min at the latest prior landing. Today's ATC-systems do not systematically provide adequate traffic information necessary for exploiting the full potential of improved capacity utilization. Under these conditions an air traffic controller cannot efficiently consider all factors of traffic flow optimization without increasing his workload. Especially with increasing traffic volumes, it becomes less likely that the potential for traffic flow optimization is recognised by ATC.

Consequently, a system support tool is needed combining information on specific flight data as well as traffic demand and calculating an optimized solution for allocating each specific flight respectively. This system shall help minimizing arrival and departure delays, balance traffic demand and reduce taxi times on the ground. The potential for such a system is certainly depending on local airport infrastructure as well as air space and traffic conditions. Within this paper first a system design is developed and then applied to the specifications of Berlin Brandenburg International Airport (BBI) including the following conditions.

- The expected traffic pattern for BBI reveals 55% of flights originating from the south and 45% originating from the north, resulting in a disparity of about 100 movements per day between the two runways. This leads to high loads and so potential arrival or departure delays on the southern runway and unused capacity on the northern runway.
- The airport layout has two independent parallel runways and a midfield concept. Parking positions near the terminal building have partly been designated to the major airlines using this airport. This creates advantages and disadvantages concerning taxi routes for certain flights.
- Due to cargo and military facilities on the north side of the airport, runway crossings become necessary for cargo and military flights arriving and departing

on the southern runway, and the risk of runway incursions may increase.

Concerning the findings from Munich Airport, it is assumed that operational ATC structures provide flexibility to introduce dynamic runway allocation procedures that help optimizing runway utilisation. However, operational constraints have to be considered.

C. Operational Constraints

Due to the operational requirements of departure route design and ATC responsibilities, a dynamic runway allocation for departure traffic is not taken into consideration for the present study: Independent simultaneous departures from parallel runways require a strict separation of utilised departure routes in order to leave the TMA geographically towards the destination airport without conflicts. A takeoff from an alternate runway requires more coordination and may lead to increased workload and reduced capacity. As such, this aspect must be considered in future research.

The predominant constraint for dynamic runway allocation of arrival traffic is the maximum approach sector capacity, which is mainly limited by the controller workload. Controller workload is mostly determined by the number of flights, the traffic-mix, and traffic activities within the sector (descent, climb, and cruise). On top, conflicting traffic situations lead to a significant increase in workload [10]. As such, the amount of deviating runway allocations is limited due to potential traffic conflicts within the responsibility of the Feeder-Controller. Thereby runway allocation requires adequate communication procedures with the cockpit crew, so that procedural adaptations have to be completed at least 10 to 15 minutes prior landing.

Finally it is emphasised that a system support tool is not fully automated so that the controller shall remain responsible for the runway allocation

III. METHODOLOGY

A significant amount of research has been performed by others in the field of runway capacity optimisation, mostly motivated by preventing traffic congestion and delays. Operations research models such as integer, linear or dynamic programming can be used for optimal allocation of interdependent arrival and departure runway system capacity to expected demand. In [11] Mixed Integer Linear Programming (MILP) was applied for optimising the allocation of flights to multiple runways over a period of one year. Apart from total delays, a multi-objective function minimises the external risk and aircraft noise to the environment. However, this concept is balancing the amount of traffic on a strategic level without focusing on the actual allocation of individual flights. A heuristic concept in [12] is allocating runway capacity in a way closely reflecting procedures of ATFM operators which makes it easier for controllers to comprehend and implement. Although this concept is based on a tactical level, it is not suited for selecting the best choice for a specific flight on the operational level. Also other factors affecting traffic efficiency, such as taxi times, schedule delay and

environmental costs should be considered for individual flights.

Similarly to the concept described in [12], a heuristic algorithm is used in this study to evaluate an optimised runway allocation under actual traffic conditions for each specific flight without considering the result of previous allocations. This implies the assumption that any locally optimised solution produces a global benefit, without inevitably reaching the optimum solution.

Arrival flights are allocated to the runway about 35 NM to 40 NM (app. 15 min) prior landing. The algorithm compares all options for runway allocation by accumulating a range of criteria for each runway. The criteria are measured in time units and valued with specific cost factors, which are based on previous studies on aircraft operating costs and airline delay costs ([13]; [14]). Results have been adapted for this particular study with regard to aircraft category, the relevant operating phase and the magnitude of delay. All cost factors represent marginal direct operating costs for an additional time unit of flight operation or delay, i.e. fuel, maintenance, crew and a limited amount of passenger compensation costs. In order to incur the costs of environmental pollution, CO₂-Emissions are calculated based on fuel consumption and valued according to the price of EU-Allowances for CO₂-Emissions (EUA) under the European Emission Trading Scheme (ETS), which will be launched for air traffic in 2012 [15].

A. Evaluation Criteria

For an arrival flight all factors are accumulated concerning operating costs during approach and landing as well as the consequences of this arrival flight on delay costs of departure traffic. This is repeated for all runway allocation options after which the option with minimum total costs is chosen. In this way the individual flight execution and the overall traffic flow with reduced arrival and departure delays can be improved. Fig. 3 illustrates the flow chart of the described algorithm.

The following criteria are evaluated: First, the flight distance is predicted from the point of optimization to the runway threshold without considering any traffic interdependencies.

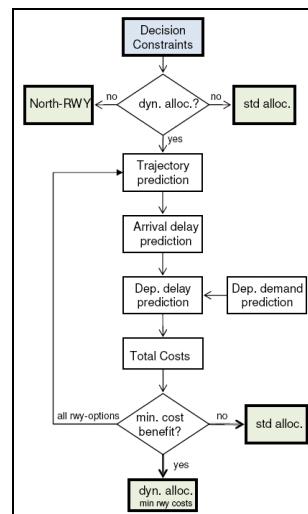


Figure 3. Flow chart of the runway allocation algorithm

This requires the knowledge of a median transition length as it is similarly used in Arrival Management (AMAN) systems. With a pre-defined speed profile a minimum flight time can be calculated and valued with a corresponding cost factor. The cost factor includes marginal operating costs for the specific aircraft category, mainly consisting of fuel and maintenance costs. Since a comprehensive approach of estimating operating costs was already undertaken by the Westminster University of London, commissioned by the Eurocontrol Performance Review Unit [14], this study refers to these results.

Second, a landing sequence is estimated based on the preceding traffic in order to predict potential arrival delays. The additional flight time is valued with a higher cost factor that also takes a low level of extra costs for flight crew and passenger compensation into account.

Third, the taxi-in time resulting from the distance between runway exit and the parking position is calculated and valued with a cost factor for aircraft ground operations. This includes a low level of fuel and maintenance costs. For reasons of simplification the taxi time does not include any traffic conflicts and waiting time on the ground.

Finally, the on-block-time can be predicted by accumulating the preceding processes. If the on-block-time deviates more than 15 min from the Scheduled Time of Arrival (STA), the resulting delay is valued additionally with a higher cost factor for crew costs and passenger compensation.

The last decision criterion values the consequences of a runway allocation for a specific flight by predicting the possible delay on departure traffic that is holding short of the relevant runway and is required to wait for runway clearance from the preceding landing. In that way the algorithm balances the costs that a specific flight induces to the traffic, taking into consideration arrival and departure demand. This requires a highly precise prediction of departure demand at the time of arrival of the allocated flight, as well as a realistic estimation of the expected delay. The delay time is then multiplied with a cost factor which includes operating costs on the ground and costs for crew and passenger compensation.

B. Delay Prediction

Delay prediction is based on the traffic demand and the required separation between two successive aircraft movements. These minima derive from radar- or wake turbulence separation minima [8]. For arrival delay determination, the predicted arrival time of two successive arrivals is compared with the required separation minima of following flights. If the time difference is less than the required separation, the subsequent flight must be delayed accordingly. This additional delay time is then used for arrival delay cost determination.

Departure delay prediction is based on a departure sequence which provides an estimated time of departure for each flight. The delay calculation for flights that would be affected by an additional landing derives from the capacity model by Newell [15]. It is assumed that arrivals are always prioritised over departures. The departure rate is therefore a function of the arrival rate. This function provides the minimum time separation between two successive departures. The resulting

departure rate is calculated with minimum time separation between two successive aircraft movements which is mostly depending on the aircraft category. This model uses a median traffic mix which is specific for BBI. Variations in wind conditions and final approach speeds are yet not considered.

C. Decision Tree Constraints

A range of constraints triggers the algorithm.

- All flights with designated parking positions on the north side of the northern runway will be allocated to the northern runway in order to avoid runway crossings for arrival traffic when taxiing.
- Arrival traffic is excluded from dynamic allocation when the limit of controller workload is reached. A simplified workload model is implemented by allowing a maximum of 10 aircraft within the sector for dynamic runway allocation at the same time. Any additional flight is allocated to the runway according to its geographical origin.
- The calculated cost difference between two runway options must exceed a minimum cost benefit to initiate a runway re-allocation (a minimum gain requirement).

IV. RESULTS

The algorithm was implemented and tested with a JAVA-based fast-time simulation, developed at the Chair of Air Transport Technology and Logistics at Technische Universität Dresden. The traffic environment was designed according to the future infrastructure and a given traffic forecast for BBI in 2011. The results of two scenarios were compared. At first a static reference scenario allocated the arrival traffic according to its geographical origin. The second simulation run applied the algorithm for dynamic runway allocation to assess the potential benefits. TAB. II summarises the main findings for both scenarios.

TABLE I. COMPARISON BETWEEN REFERENCE AND DYNAMIC SCENARIO

	scenarios:		reference	dynamic
	abs.	rel.	abs.	rel.
aircraft movements (sum)				
north-rwy (25R)	509	43%	596	50%
south-rwy (25L)	678	57%	591	50%
re-allocations				
(abs. and share of total arrivals)	0	0%	249	42%
delays (sum and relativ reduction)				
Arrival (25L/R)	10:12:43		4:02:02	-60%
Departure (25L/R)	25:13:40		16:11:22	-36%
taxis-in time (sum and rel. reduction)	46:51:23		42:28:35	-9%
main-apron crossings (sum and rel. red.)	157		118	-25%

The analysis of the simulation results reveals that dynamic runway allocation creates an equalised traffic distribution on both runways. The unbalance between the two runways of 100 movements per day is resolved. Major peaks with more than 50 movements per hour on the southern runway are reduced to a maximum of 47 movements per hour (see Fig. 4). However, the arrival-departure ratio shifted from 50-50 % to 57-43 % with a higher share of arrivals on the northern runway and

more departures on the southern runway. Especially during periods of extensive departure demand on one runway, arrivals are allocated to the alternate runway (i.e. departure peak at 6:00).

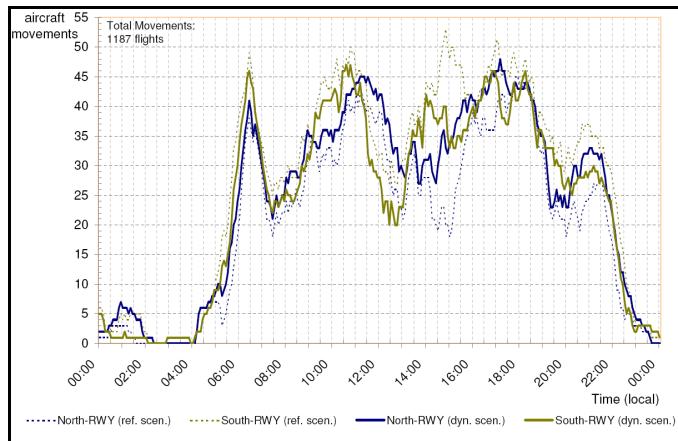


Figure 4. Comparison of total movements as rolling hour for North and South runway

Fig. 5 illustrates the accumulated delay of arrival and departure traffic on both runways. Overall delays can be reduced on the northern runway by 13.6 % (-1.8 h in total per day) and on the southern runway by 60.4 % (-13.4 h in total per day). On average, arrival delays are reduced by 37 s per flight. Delays for departure flights are reduced by 55 s per flight. Especially during times of equally high arrival and departure demand (i.e. 9:00 to 12:00 and 16:00 to 18:00) delays can be reduced significantly.

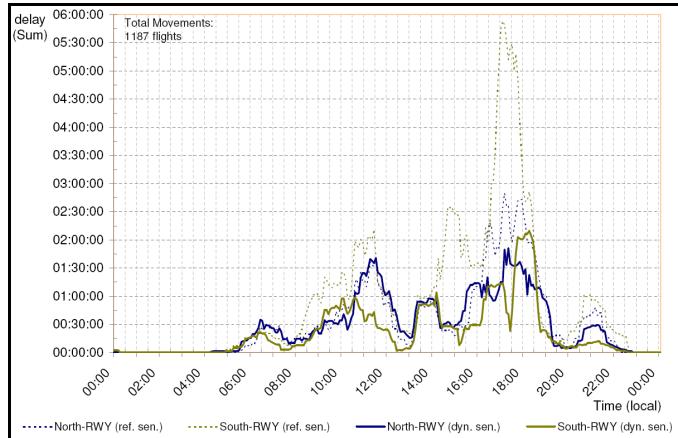


Figure 5. Sum of delay in floating hour for North and South runway with comparison of both scenarios.

Taxi times are reduced by 9.4 % in total (-4.4 h per day) which corresponds with a reduced taxi time of 26 s on average per flight. Extremely high individual taxi times in the static scenario (up to 12 minutes) can be resolved completely. The maximum recorded taxi-in time during the dynamic allocation scenario is 8.5 minutes. For some flights taxi-in times are reduced by more than 8 min, whereas some flights have extended taxi-in times by 1 min to 5 min. This is due to the

integration of all decision criteria in one optimisation function where one criterion might overvalue other criteria. In this case extended taxi-in times might be accepted for reduced arrival delays or resolved departure queues.

The re-allocation of arrivals for reasons of reduced taxi time has direct consequences on the complexity of ground traffic and potential conflicts. Due to the airport layout of BBI aircraft taxi routes cross the main apron when landing on the south runway but parking on the northern apron and vice versa. The usage of these taxi route relations is reduced by 25 %.

Fig. 6 illustrates the arrival traffic flow, the share of re-allocation (orange line) and the distribution to both runways (blue and green line) over 24 h. In total a share of 42 % of all arrivals (249 flights) is re-allocated to the opposite runway.

In 43 cases the workload limit of the Feeder-Controller prevented a re-allocation (see red line in Fig 5). Additionally, 33 flights did not exceed the minimum cost benefit limit, so that these flights remained on the standard approach. Especially in periods of high arrival demand, the potential of airspace conflicts increases because of crossing traffic.

The reduction in arrival and departure delay as well as taxi times leads to a reduction in fuel burn of app. 19 tonnes. This leads to potential savings of 8.143 € for airlines and about 60 t less CO₂-Emissions.

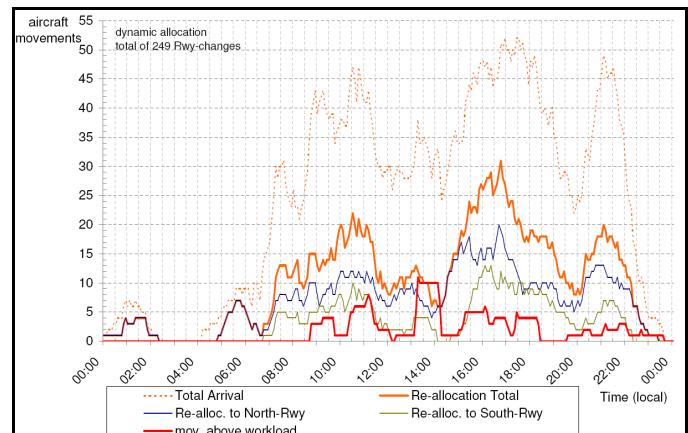


Figure 6. Arrival movements with Re-allocation to North and South runway and blocked allocations due to workload limitations.

V. OPERATIONAL IMPLEMENTATION

According to the objectives of SESAR and considering the recommendation to establish a centralised and demand-orientated resource-management at airports, it is advised to implement a dynamic runway allocation system at an Airport Operation Center (AOPC) under the responsibility of the airport authority in close partnership with the local ATC-Tower services. The core principle for operational decision making within an AOPC environment is Airport Collaborative Decision Making (A-CDM) process which provides a structured and frequent data exchange between airport operators, Air Navigation Service Providers (ANSP), airlines and handling agents. Access to updated operational flight plan

data is a vital requirement for reliable trajectory and delay prediction.

Traffic predictions and workload models are best done by ATC-services and provided as input data for the algorithm. The final calculations for runway allocations should be done according to the CDM philosophy within the APOC and sent to the operational ATC-controller via an appropriate data-link. Other research projects have identified practical means to present the suggestion for runway allocation to the controller so that the information is easy to comprehend and does not disturb basic controller tasks. In [17] it is suggested to include the allocation information within the aircraft label on the radar screen of the approach controller. The system should allow the user to interact and accept or disregard the information.

This concept for dynamic runway allocation is based on operational procedures used by ATC and flight crews. The system support is designed in a way that the controller is not limited in his current work tasks. Any runway advice can be rejected or ignored without any consequences for traffic safety. Compared to subjective evaluation by the controller, the system support is clearly decreasing workload by providing a runway suggestion that includes a range of criteria by combining data from various sources. Without a system support this is too much information for a human being to be able to process at a given time especially with increasing traffic volumes. The harmonised traffic flow also reduces the coordination between Tower and Approach Control, because the allocation of arrival traffic already considers departure traffic demand. However, a potential increase in conflicting airspace situations, and the necessity for more radio communication, can limit the usability of dynamic runway allocation under high traffic demand.

VI. CONCLUSION AND FURTHER RESEARCH

The designed algorithm in this paper is based on the integration of several criteria in a total cost function which is used for the evaluation of runway options for specific arrival flights. The architecture of the algorithm is sequential and not iterative so that a total optimum for the entire arrival and departure traffic is not inevitably achieved. A sequential optimisation provides the necessary foundation for further research. However, with this method it is possible to control and allocate arrival traffic according to the actual traffic condition and capacity situation and to create overall improvements in the traffic flow.

The implementation within a simulation tool showed clear benefits for the traffic flow at BBI. The utilisation of both runways was well balanced and delays were reduced by 55 seconds per departure and 37 seconds per arrival. Taxi-in times were reduced by 26 seconds per arrival on average. Crossings of the main apron, which can produce potential ground conflicts, were cut by 39 movements. This leads to potential savings of more than 8.000 € in fuel costs and 60 t of CO₂-Emissions per day.

A range of improvements for the algorithm was identified. Data analysis has shown that under high traffic volumes the predicted departure sequence is not stable enough and

provides an over or under estimation of departure delays. This has direct consequences on the reliability of the departure delay prediction which is an important criterion for the runway allocation. Therefore, the departure sequence should be updated frequently in order to take into account the actual rate of arrivals.

Further improvements could be made by introducing a limited iteration process for the runway allocation in order to compare the cost benefit of the preceding and consecutive flight, and to adapt the runway allocation if necessary. However, it is vital that only one final suggestion for runway allocation is presented to the ATC-controller well in advance, so that the controller and the flight crew are not confused by changing allocations. For this purpose, the internal calculation process should begin at least 30 min or 100 NM and stop latest 10 min or 30 NM prior landing.

Additionally, the workload model used in this algorithm should be extended so to consider potential airspace conflicts as well. This requires a precise trajectory prediction and sequence planning for arrival traffic.

For implementation purposes further research must be done on data exchange and technical data-link solutions between the APOC, ATC-controllers and the cockpit crews. Special attention should be drawn to an appropriate human-machine-interface for the controller. A collaborative development and operation of such a system is best done on the basis of a common decision-making platform with all partners involved.

REFERENCES

- [1] SESAR Consortium, *The performance target D2*, Eurocontrol, December 2006.
- [2] Episode 3 Consortium, *Runway management E1 D2.2-034*, Version 1.0, European Commission, January 2009.
- [3] SESAR Consortium, *The ATM target concept D3*, Eurocontrol, September 2007.
- [4] Günther, T., Fricke, H., *Potential of speed control on flight efficiency*, in Proceedings of the 2nd International Conference on Research in Air Transportation, Belgrade, June 2006.
- [5] Günther, T., Hildebrandt, M., Fricke, H., Strasser, M., *Contributions of advanced taxi time calculation to airport operations efficiency*, Air Transport and Operations Symposium 2010, in press.
- [6] International Civil Aviation Organization, *Annex 14, aerodromes*, vol. 1, 4th ed., Chapter 3, November 2004.
- [7] International Civil Aviation Organization, *Manual on simultaneous operations on parallel or near-parallel instrument runways (SOIR)*, Doc 9643, 1st ed., Appendix, 2004.
- [8] Deutsche Flugsicherung GmbH, *Betriebsanweisung Flugverkehrskontrolle (BA-FVK)*, DFS, Langen, March 2007.
- [9] Deutsche Flugsicherung GmbH, *Aeronautical information publication Germany – Luftfahrtshandbuch Deutschland, aerodromes (AIP-AD)*, DFS, September 2006.
- [10] Mensen, H., *Moderne Flugsicherung, Organisation, Verfahren, Technik*, 3rd ed., Springer Verlag, Berlin, 2004.
- [11] Heblrij, S., Wijnen, R., *Development of a runway allocation optimisation model for airport strategic planning*, Transportation Planning and Technology, vol. 31, no. 2, pp. 201-214, April 2008.
- [12] Janic, M., *A heuristic algorithm for the allocation of airport runway system capacity*, Transportation Planning and Technology, vol. 30, no. 5, pp. 501-520, October 2007.
- [13] International Civil Aviation Organization, *Cost tables of CNS/ATM planning and evaluation tools, base-line aircraft operating costs*

- (appendix), 4th Meeting of the ALLPIRG/Advisory Group, February 2001.
- [14] Transport Studies Group, *Evaluating the true cost to airlines of one minute of airborne or ground delay*, University of Westminster, May 2004.
- [15] European Environment Agency, *Air pollutant emission inventory guidebook*, Technical Report no. 6, 2009.
- [16] Newell, G. F., *Airport capacity and delays*, Transportation Science, vol. 13, no. 3, pp. 201-241, 1979.
- [17] Lee, K., Sanford, B., *The passive final approach spacing tool (pFAST), human factors operational assessment*, 2nd USA/Europe ATM R&D Seminar, Dec. 1-4 1998, Orlando, 1998.

AUTHOR BIOGRAPHIES

Martin Fritzsche studied traffic engineering at Technische Universität (TU) Dresden where he received his Diploma in 2010. In his study he worked on different projects regarding ATM and environmental sustainability of air traffic. The present investigation is part of his diploma thesis supervised by TU Dresden and Flughafen Berlin-Schönefeld GmbH between 2009 and 2010.

Thomas Günther works at Technische Universität Dresden, Chair of Air Transport Technology and Logistics as a scientific assistant. Within his PhD thesis he is currently working on the assessment of efficiency improvements under consideration of ATM network effects. The development and application of a proper methodology shall contribute to a better understanding of according potentials regarding lateral, vertical and speed profiles as well as queue management and surface movement aspects. Thomas studied traffic engineering at TU Dresden where he received his Diploma in 2004.

Hartmut Fricke studied Aeronautics and Astronautics at Technische Universität Berlin from 1985–1991. From 1991 to 1995 he was a research fellow in Flight Operations, Airport Planning, and ATM at TU Berlin, where he completed his doctor thesis in ATM (ATC-ATFM Interface). In 2001 he finished his Habilitation on “Integrated Collision Risk Modeling for airborne and ground based systems”. This included HIL experiments with an A340 full flight simulator in co-operation with EUROCONTROL Experimental Centre (EEC). Since December 2001 he has been Head of the Institute of Logistics and Aviation, and Professor for Aviation Technologies and Logistics at TU Dresden. In 2006 he was appointed Member of the Scientific Advisory “Board of Advisors” to the Federal Minister of Transport, Building and Urban Affairs in Germany.

Track 4
CNS/ATM

An Analysis of Delays in Air Transport in Japan

Kota Kageyama

Electronic Navigation Research Institute
7-42-23, Jindaiji-higashi-machi, Chofu
Tokyo, Japan
Email: kage@enri.go.jp

Yutaka Fukuda

Electronic Navigation Research Institute
7-42-23, Jindaiji-higashi-machi, Chofu
Tokyo, Japan
Email: fukuda@enri.go.jp

Abstract—To cope with the increase in air traffic demand, improving ATM (air traffic management) performance is important. This paper describes an analysis of delays in air transport in Japan as part of ATM performance evaluations. The study examined arrival punctuality and departure punctuality at major Japanese airports. Punctuality is measured according to scheduled times. Characteristics of punctuality in Japan are compared to those in the United States and in Europe. Delay is studied in terms of conformity with flight-plans. It is assumed that high conformity is represented by a small average and distribution of delay. To study the conformity in each operational phase, an aircraft operation is divided into four distinct phases: pre-departure, taxi-out, airborne and taxi-in. Delays are calculated for standard times and the averages and the standard deviation is studied for each phase. The division into operational phases revealed that pre-departure delay is the main driver of fluctuation in delay. We also examined ATFM (air traffic flow management) impact on pre-departure delay.

Index Terms—ATM Performance, delay, punctuality, air traffic flow management

I. INTRODUCTION

ATM (air traffic management) is the dynamic and integrated management of air traffic and airspace through the provision of facilities and seamless services to airspace users in collaboration with all involved stakeholders, with the objective of achieving safe, economical and efficient operations. To accommodate increase in air traffic demand, ATM performance has been significantly improved in the last few decades. However, since increased air traffic demand is anticipated, further ATM performance improvements are required. For that purpose, ATM performance evaluations are required to provide valuable assessment information. Through such ATM performance assessment, performance bottlenecks can be identified and prioritized in order to be dealt with and hopefully removed appropriately. Furthermore, any ATM performance study should facilitate the estimation of the effect of planned improvements prior to implementation.

Since ATM has, by definition, multiple objectives to accomplish, its performance must be examined through multiple assessment viewpoints. The ICAO (International Civil Aviation Organization) has defined KPA (key performance areas)[1]. The KPA are comprised of 11 areas corresponding to social impact (safety, security, environment), ATM prosperity (access and equity, participation by the ATM community), and ATM operation performance (cost effectiveness, capacity, efficiency,

flexibility and predictability). The KPA are by nature almost contradictory[2].

Amongst many performance metrics, the study focuses on aircraft operational delays. Delay relates to the area of punctuality and predictability[3]. In the interest of financial performance and predictability of operations, it is usually deemed desirable that the actual arrival times agree with their scheduled values. However, it was realized that, due to the application of ATM procedures, as well as due to other contributing factors such as weather or carrier-action, the actual arrival times at destination airports usually differ from the scheduled arrival times. A numeric target for delay reduction is set out in the ATM transformation plan presented in [4]. At the same time, current delays are analyzed at some locations. For instance, in Europe, CODA (Central Office for Delay Analysis) publishes aircraft operational delays on a regular basis[5].

Concerning air transport as well as ATM, each location has its specific character. Thus, delays need to be studied for each location. This paper examines delays in Japan. Firstly, it compares punctuality at Japanese major airports. Punctuality is measured based on industry-standard indicators: the percentage of arrivals/departures more than 15 minutes later than the scheduled time. Delay is also studied in terms of conformity with flight-plans. The conformity is measured based on delay from flight-plans. It is assumed that a small average and distribution of delay achieve high conformity. In the conformity study, an aircraft flight operation is divided into distinct phases based on values of the time stamps of various events during an aircraft's operational mission. Averages and the standard deviation of delay were calculated for each phase. In addition, the paper looks at ATFM (air traffic flow management) impact on pre-departure delay.

II. PUNCTUALITY STUDY

A. Analyzed Data

The airports were chosen for this analysis based on domestic flights movements. According to statistical data, the top five airports for domestic flights movements were Tokyo International (Haneda) (RJTT), Osaka (RJO0), Fukuoka (RJFF), Naha (ROAH) and Sapporo (RJCC) in the year of 2007. Figure 1 shows the domestic landing count comparison during the year. The total landing count at the airports accounted for around 38% of IFR traffic volume at Japanese airports.

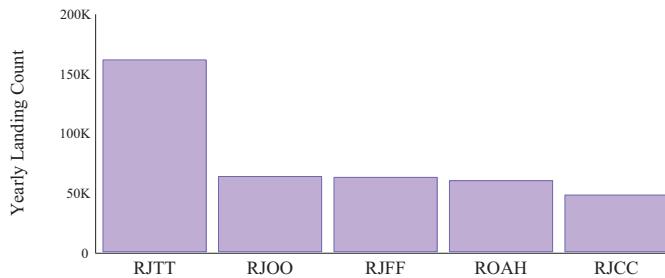


Fig. 1. A Comparison of Landing Count in 2007

Amongst the five airports, RJTT at which the landing count was the highest, plays the role of a domestic hub airport. Although international flights landed at the airports (RJTT:2%, RJOO:0%, RJFF:11%, ROAH:2%, RJCC:2% of the entire volume), due to data availability, only domestic flights were covered in this analysis instance.

The generally accepted indicator for air transport delay is the percentage of arrivals more than 15 minutes later than the scheduled time[3]. Arrival punctuality was calculated as the percentages of arrivals 15 minutes late or less. Likewise, the percentages of departures no more than 15 minutes later than the scheduled time were computed as departure punctuality.

The analyzed data were recorded in February, June, August, October, December of 2007, in April, June, August, October, December of 2008, and in February 2009. 6-7 days worth of data were gathered for each of these months. In total, 76 days worth of data were analyzed. The following data items were obtained from ATM system journals: actual gate-in times (ABIT: actual block-in time) and gate-out times (AOBT: actual off-block time) were obtained from SMAP (spot management and planning system) journals. Scheduled times were obtained from timetables. If necessary, radar data from RDP (radar data processing system) journals were used to study the traffic situation.

For delay analysis, there are external factors influencing. Bad weather conditions are major examples of such factors. For instance, snowfall or strong wind can cause runway closure. To measure ATM performance exclusively, data under these conditions should be ignored. However, due to limitations in weather data sources, all the gathered data were analyzed regardless of weather and other conditions.

B. Analysis Results

Figure 2 shows the monthly punctuality at the five airports combined. The red line represents the arrivals punctuality, and the blue line represents the departure punctuality.

The indicators were also calculated for each of the five airports. Figures 3, 4, 5, 6, and 7 show the monthly arrival and departure punctuality at each airport.

At RJTT (Figure 3), the arrival punctuality tended to be lower than at the other airports. At RJOO (Figure 4) and RJFF (Figure 5), the punctuality was relatively high.

At ROAH (Figure 6), the arrival and departure punctuality demonstrated significant increases and decreases. On the

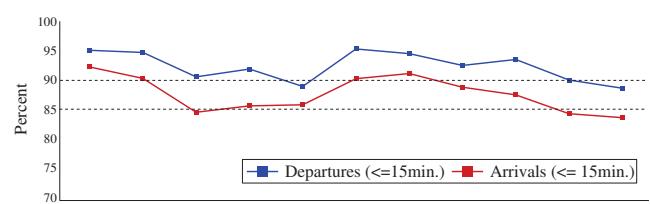


Fig. 2. Air Transport Punctuality (To/From the Five Airports)

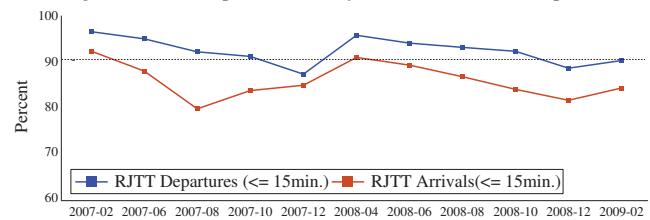


Fig. 3. Air Transport Punctuality (To/From RJTT)

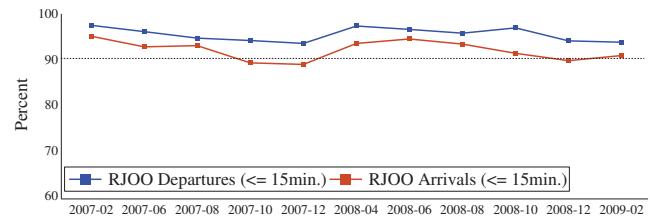


Fig. 4. Air Transport Punctuality (To/From RJOO)

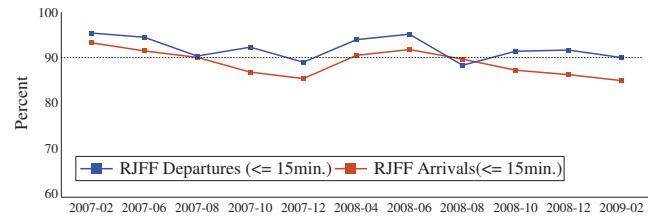


Fig. 5. Air Transport Punctuality (To/From RJFF)

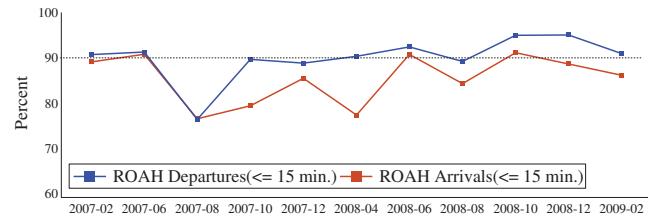


Fig. 6. Air Transport Punctuality (To/From ROAH)

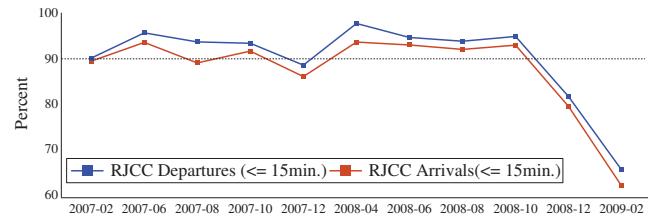


Fig. 7. Air Transport Punctuality (To/From RJCC)

whole, punctuality at ROAH was worse than at the other airports. The reasons for this fluctuation need to be examined in the future study.

At RJCC (Figure 7), arrival and departure punctuality dropped significantly in December 2008 and February 2009. Regional peculiarities in Japan are important here. Because Japan extends to the north and south, the weather conditions vary amongst areas. RJCC is located in the north. As a result, runway-closure due to heavy snow can occur during winter at this airport while snow rarely falls at the other four airports. The RDP journal implies the possibility of runway-closure during the winter months. Except for these months, punctuality at RJCC was almost the same level as at RJOO and RJFF.

C. Discussion

Overall, arrival punctuality was 87.6% and departure punctuality was 92.3%. The same indicators were calculated for the main airports in Europe and the United States. In 2007, arrival punctuality was around 78% in Europe and 76% in the United States, and departure punctuality was around 78% in Europe and 80% in the United States[6].

The selection criteria for data (e.g. airports, time periods, etc.) in this study were different from the study for the United States and Europe[6]. The study for the United States and Europe covered the main 34 airports in the United States (OEP34) and Europe. Traffic to and from the main 34 airports represented some 69% of all the IFR flights in Europe and 64% in the United States[6]. In this study, the chosen airports represented around 38% of IFR traffic volume at Japanese airports. Although more airports should be chosen for strict comparison, the chosen airports here were outstanding in terms of domestic traffic volume and passenger enplanements. For instance, the airports accounted for around 64 % of all domestic passenger enplanements.

In terms of data amount, this study covered 76 days, which was around 10% of the corresponding period (2 years). More data need to be analyzed for a detailed study.

In [6], it was also pointed out that the gap between departure and arrival punctuality was almost nil in Europe whereas it was significant in the United States. From Figure 2, it was observed that Japanese gap was more similar to that of the United States.

There was a similar trend because increases and decreases between arrival and departure punctuality were the same. The trend implied the possibility of reactionary departure delay caused by arrival delay. To study the possibility, the connection between late departures and late arrivals need to be examined in the future studies.

III. CONFORMITY STUDY

A. Definition of the Phases

This section presents the analysis results for the conformity with flight-plans. To achieve high conformity, delays in each flight operation should be small, i.e. delays should average zero. In addition, the study must focus on distribution. High averages with a small distribution implies a tendency around a

point far away from zero. In this case, there is the possibility that the method of delay measurement requires refinement. As a result, the study look at averages and the standard deviation of the delays.

To examine the conformity in detail, the entire operation between gate-out and gate-in was divided into distinct phases and the conformity was studied in each phase. The recorded times of flight operation events (gate-out, take-off, touch-down and gate-in) were taken as the standard and precisely-defined breakpoints of the various phases[7]. In this study, based on the recorded times of the various events, the flight operation was classified into the following four phases:

- 1) Pre-departure (ending at gate-out),
- 2) Taxi-out (beginning at gate-out and ending at take-off),
- 3) Airborne (beginning at take-off and ending at touch-down),
- 4) Taxi-in (beginning at touch-down and ending at gate-in).

The following data items for the calculation was obtained from ATM system journals: the flight plan data items were gathered from FDMS (flight data management system) journals. Take-off and touch-down times were also taken from FDMS journals. As was the case with punctuality, actual gate-in times and gate-out times were obtained from SMAP journals.

Figure 8 shows the idea of the phase classification. The conformity indicators were defined as the variance between planned and actual times and calculated as follows.

B. Calculation Methods

1) *Pre-departure*: The pre-departure delay was equivalent to the variance between the actual and scheduled gate-out time. The actual gate-out times were recorded as AOBT. The scheduled gate-out times are STD (scheduled time of departure).

In the flight-plan, planned departure time is recorded as EOBT (estimated off-block times). STD are initially used as EOBT. However, in the case of scheduled departure times being delayed, flight-plans are re-issued and EOBT are renewed. Although it can be measured as the variance between AOBT and EOBT, pre-departure delay was regarded as the variance between initial scheduled times and actual time and calculated as:

$$AOBT - STD. \quad (1)$$

The ATM-related factors in the pre-departure delays included the following items.

- Airport surface design,
- Runway/taxiway congestion,
- Convergence of departures using the same SID (standard instrument departure) route,
- Take-off time adjustment by ATFM (air traffic flow management).

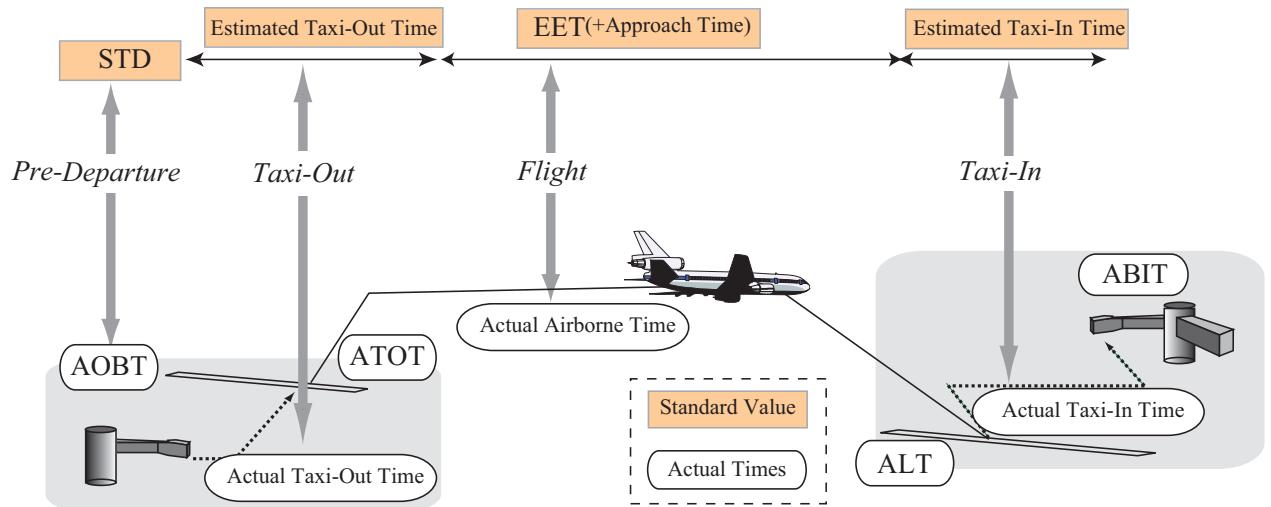


Fig. 8. The Phases

2) *Taxi-out:* Actual taxi-out time was defined as time from AOBT to ATOT (actual take-off time). To measure delay, standard taxi-out time T_o needs to be defined for every pair of departure-gates and take-off runways. The taxi-out delay was calculated as:

$$(ATOT - AOBT) - T_o. \quad (2)$$

The ATM-related factors in the taxi-out delay include airport surface design and runway/taxiway congestion.

The ATFM system had the data set of estimated time between each pair of departure gates and take-off runways. The data set was used for estimating the take-off time. For convenience, the data set were used as the tentative standard. It should be noted that the data set were contained representative values and were not always equivalent to the shortest (unimpeded) values.

3) *Airborne:* Prior to departure, airlines estimate the airborne time as EET (estimated elapsed time) based on planned routes, and it is recorded in the flight-plan. On the other hand, the actual value can be calculated as time from ATOT to ALT (actual landing time). As a result, the airborne delay is calculated as:

$$(ALT - ATOT) - EET. \quad (3)$$

The ATM-related factors in airborne delay include airspace congestion.

EET corresponded to the approximate time from take-off to approach commencement. Time from approach commencement to touch-down was uniformly set to 5 (five) minutes and added to the EET.

For the calculation of EET, the wind-speed and runway direction predictions were taken into consideration. Since wind-speed influences the ground speed and runways direction influences the flight distance, the prediction of wind-speed and runways direction had an impact on the calculated EET. However, the prediction results may vary amongst airlines. As a result, even though the planned routes and other items were identical, EET may be different depending on the airlines. In this study, it was assumed that EET represented the airlines' expected airborne time and the discrepancy of the prediction was not taken into consideration.

4) *Taxi-in:* Actual taxi-in time was defined as the variance between ALT and ABIT (actual gate-in time). Standard taxi-in time was represented as T_i . The taxi-in delay was calculated as:

$$(ABIT - ALT) - T_i. \quad (4)$$

The ATM-related factors of the taxi-in delay include airport surface design and gate congestion.

As is the case with taxi-out, T_i has to be defined for every pair of touch-down runways and arrival gates. The data set from the ATFM system was used to define T_i . Because the ATFM system did not define taxi-in times, the pairs of departure gates and take-off runways were converted into pairs of touch-down runways and arrival gates. Deducting an estimated push-back time of three (3) minutes from the corresponding T_o , the result was used as the T_i standard value.

C. Analysis Results

1) *Averages:* The study examined the conformity of arrivals at the major airports. The airports and time periods used for

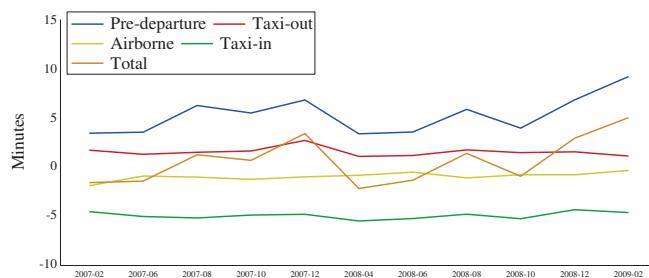


Fig. 9. Trends in the Averages of the Delays (the Five Airports, Monthly)

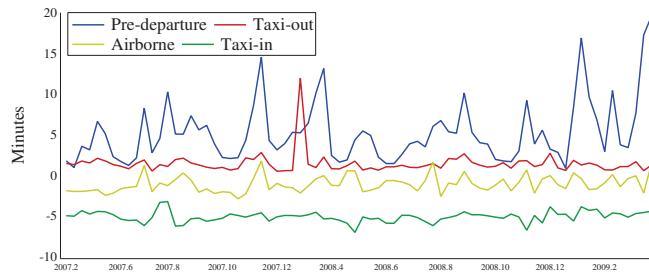


Fig. 10. Trends in the Averages of the Delays (the Five Airports, Daily)

this analysis were identical to those in the punctuality study. Due to SMAP application coverage, pre-departure delay and taxi-out delay were calculated only for the flights within at the specified airports.

Figure 9 shows the monthly averages at the five airports combined. From the figure, it was observed that pre-departure delay fluctuated and the other delays were relatively constant. In comparison, Figure 10 shows the daily averages. Comparing Figure 10 with Figure 9 indicates that the fluctuation from month to month was much smaller than from day to day. This can be attributed to anomalous events during operation. With the current technology, anomalous operation events due to unexpected weather or other occurrences are inevitable. As a result, there could be disturbances. These disturbances could affect the calculation of the daily averages to a remarkable degree. On the other hand, their effect was neutralized to some extent in the calculation of the monthly averages in which 6-7 days worth of data were combined.

Monthly averages were also calculated for arrivals at each of the airports. Figures 11, 12, 13, 14 and 15 show the monthly averages at each airport. At all the airports, pre-departure delays fluctuated more than in other delay phases. In particular, on December 2008 and February 2009, a high level of pre-departure delay was observed in RJCC arrivals. In these months, airborne and taxi-in delays for RJCC arrivals also increased. This can be attributed to the possibility of runway closure mentioned in II-B.

At the five airports, taxi-out delays were virtually constant and airborne and taxi-in delays fluctuated slightly.

2) *The Standard Deviation:* Figure 16 shows the overall monthly standard deviation at the five airports. From the figure,

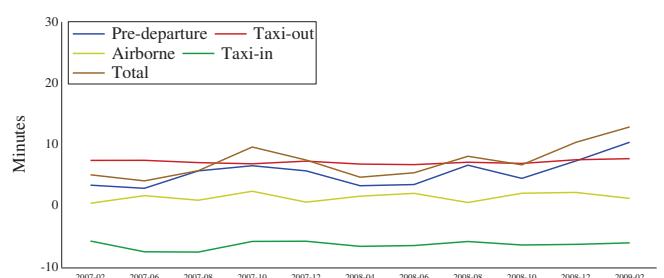


Fig. 11. Trends in the Averages of the Delays (RJTT Arrival)

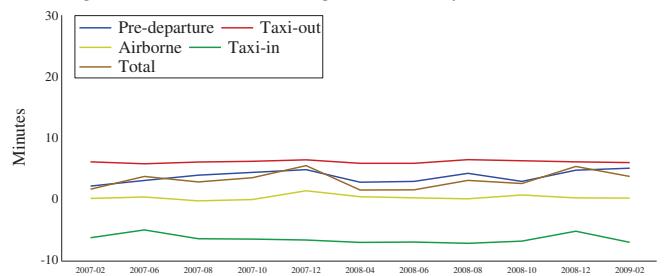


Fig. 12. Trends in the Averages of the Delays (RJOA Arrival)

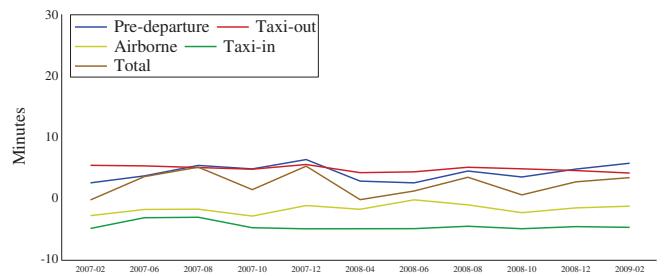


Fig. 13. Trends in the Averages of the Delays (RJFF Arrival)

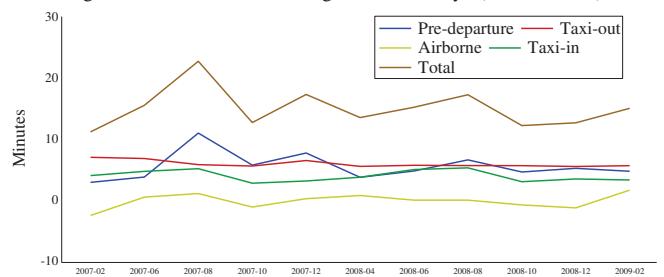


Fig. 14. Trends in the Averages of the Delays (ROAH Arrival)

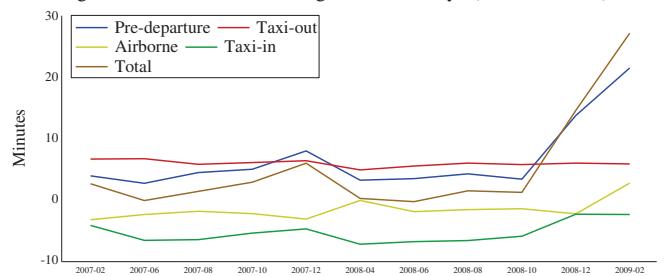


Fig. 15. Trends in the Averages of the Delays (RJCC Arrival)

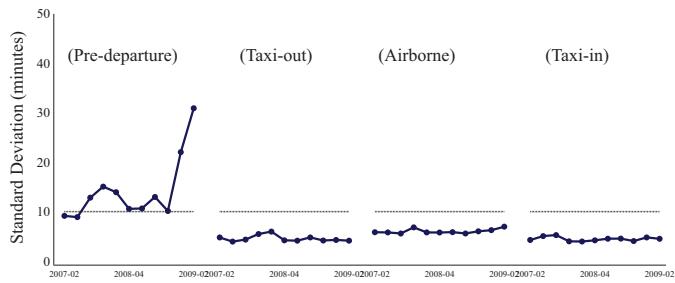


Fig. 16. Trends in the Standard Deviation of the Delays (Five Airports, Monthly)

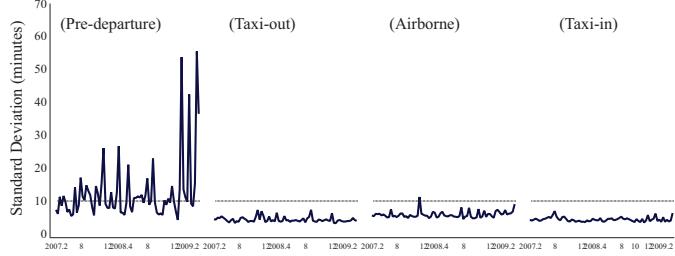


Fig. 17. Trends in the Standard Deviation of the Delays (Five Airports, Daily)

it was observed that the standard deviation of the pre-departure delay fluctuated widely from month to month. In addition, the standard deviation of the pre-departure delay was always much higher than for other delay phases. While the standard deviation of pre-departure delay was always more than 10 minutes, in other phases it was less than 10 minutes.

Figure 17 shows the overall daily standard deviation. The daily fluctuation of pre-departure delay was wider from day to day than from month to month.

Standard deviations were also calculated for each of the airports. Figures 18, 19, 20, 21 and 22 show the monthly standard deviations. As was the case with the five airports combined, the standard deviation of pre-departure delay was always much higher than in the other delay phases at each of the airports. At RJCC, the standard deviation of airborne delay and taxi-in delay increased in December 2008 and February 2009. Like the averages, this can be attributed to the possibility of runway-closure.

Except for pre-departure delay, the standard deviation of delays proved to be relatively constant. In other words, after gate-out, delay distribution was relatively constant. This can be attributed to ATM which can possibly manage delay during flight operations. On the other hand, factors outside ATM coverage such as carrier-action and reactionary-delay, i.e. pre-departure delay due to arrival delay, can often affect pre-departure delay. The large standard deviation of pre-departure delay can be attributed to these factors.

D. Discussion

Table I shows the averages and the standard deviation of the delays for all the data. The table demonstrates that the averages

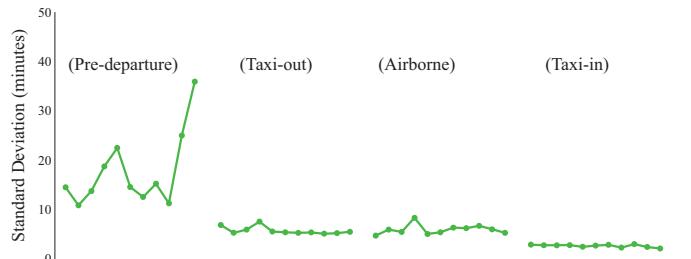


Fig. 18. Trends in the Standard Deviation of the Delays (RJTT Arrival)

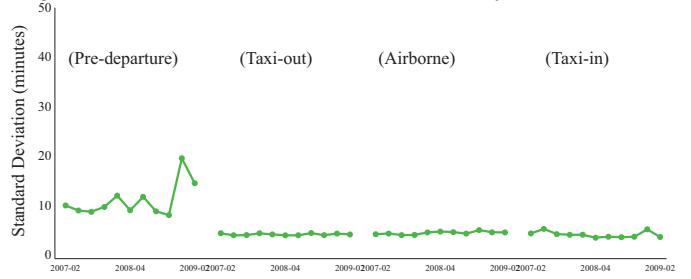


Fig. 19. Trends in the Standard Deviation of the Delays (RJOO Arrival)

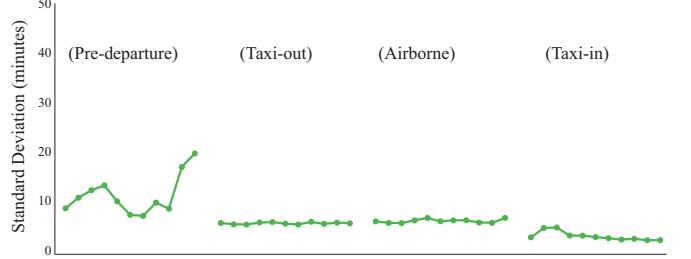


Fig. 20. Trends in the Standard Deviation of the Delays (RJFF Arrival)

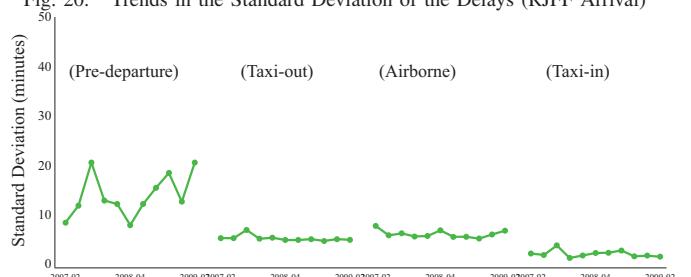


Fig. 21. Trends in the Standard Deviation of the Delays (ROAH Arrival)

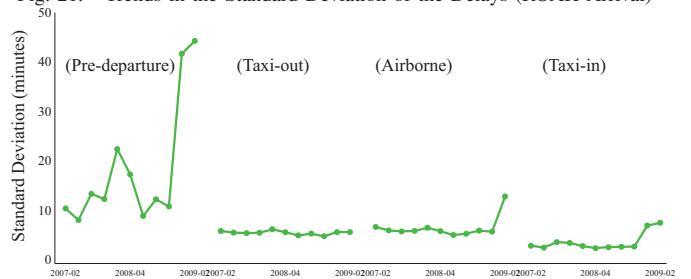


Fig. 22. Trends in the Standard Deviation of the Delays (RJCC Arrival)

TABLE I
THE AVERAGES AND THE STANDARD DEVIATION (MINUTES)

Phase	Average	Standard Deviation
Pre-departure	5.2	15.4
Taxi-out	1.5	4.6
Airborne	-1.1	6.0
Taxi-in	-5.1	4.5

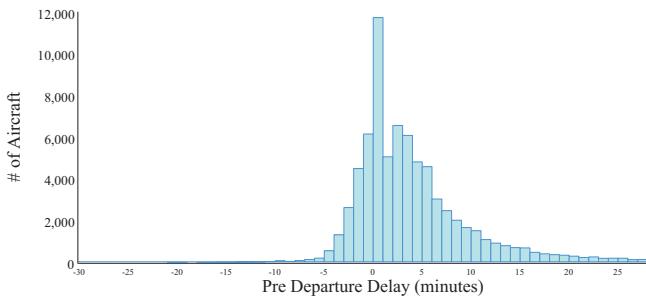


Fig. 23. Frequency Distribution : Pre-departure delay (Five Airports)

and the standard deviation of pre-departure delay were much higher than for the other delay phases.

Taxi-in delay averaged negative values. It is unreasonable to assume that the actual taxi-in time is 5 (five) minutes shorter than the expected value. The average implies that the calculation method for taxi-in delay needs to be refined.

The averages of taxi-out delay and airborne delay were close to zero. The averages of actual taxi-out time and actual airborne time were 12.8 minutes and 72.4 minutes, respectively. Taking the average of the actual time into consideration, the distribution of taxi-out delay can be regarded as being much more than the distribution of airborne delay. It is generally recognized that uncertainty in taxi-out delay reduces the conformity of the entire operation[8].

This study indicated a high-level of pre-departure delay. Pre-departure delay was calculated by examining the flights departing from five major Japanese airports. On the other hand, the punctuality study presented in II indicated that departure punctuality in the same data sets was high (92.3%).

Figure 23 shows the overall frequency distribution of pre-departure delay at the five airports. The frequency distribution indicates that although the percentages of pre-departure delays of more than 15 minutes was smaller, the distribution was spread more widely. The large distribution explains the high averages as well as the standard deviation for pre-departure delay.

As was the case with the study of flights in Europe and in the United States[6], pre-departure delay proved to be the main driver of delay fluctuation in overall flight operations.

IV. STUDY OF THE CAUSES OF PRE-DEPARTURE DELAY

In the analysis results mentioned above, pre-departure delay fluctuated widely over the observed months. The causes of the fluctuation need to be studied. For instance, CODA classifies

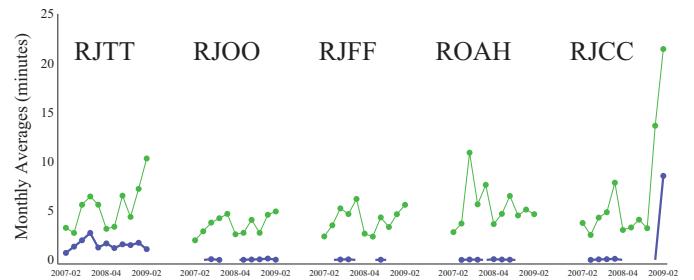


Fig. 24. Averages of Pre-Departure Delay and ATFM Delay (Monthly)

TABLE II
STATISTICAL DATA OF ATFM DELAY

	% of Flights Delayed	Delay per Flight (minutes)	Delay per Delayed Flight (minutes)	Pre-departure Delay per flight (minutes)
All	6.7%	0.7	10.5	5.2
RJTT	15.8%	1.5	9.8	5.3
RJCC	16.4%	8.5	51.8	21.4

the causes of pre-departure delay into categories based on the data supplied by airlines[5]. However, because the data on delay causes were not available, it was impossible to classify the causes of pre-departure delay in this study.

Using ATM journals, the study examined one ATM-related cause of pre-departure delay. ATFM is a typical ATM function that affects pre-departure delay. To avoid airspace congestion, ATFM uses take-off times as EDCT (expected departure clearance time) to delay flights prior to departure. As a consequence, gate-out or taxi-out time can be delayed. Meanwhile, the airborne delay should be reduced.

It was possible that the gate-out time adjustment for ATFM caused for the observed fluctuation in pre-departure delay.

To analyze the impact of ATFM, ATFM delays were calculated as:

$$EDCT - ETD. \quad (5)$$

ETD (expected departure time) which corresponded to scheduled take-off time, was calculated as the sum of EOBT and T_o . EDCT was issued only for the flights with regulated take-off times. ATFM delay was calculated for each of the EDCT-issued (delayed) flights. The total of ATFM delays were then computed.

The monthly averages of the ATFM delays were calculated by dividing this total by the number of all the arrivals in each month. It should be noted that the ATFM delay did not cover the entire ATFM adjustment, because sometimes take-off times were adjusted without an EDCT being issuance.

Figure 24 shows a comparison of the monthly averages between pre-departure and ATFM delays. The comparison represents the magnitude of ATFM impact on the pre-departure delay. In the figure, the averages were categorized by arrival airports.

Primarily RJTT arrivals incurred ATFM delay. This indicated that the frequency of demand and capacity balancing due to the heavier volume of arrival traffic at the airport.

It was also observed that in February 2009, RJCC arrivals incurred a high level of ATFM delay. As mentioned earlier, there was the possibility of runway-closure during February. The ATFM delay could be attributed to this. If this is excluded, there was little major ATFM impact on monthly pre-departure fluctuation at the airport.

Table II shows the statistical data for ATFM delay from the analysis results. The table presents the averages of pre-departure delay for comparison. The statistical data were calculated for the overall arrivals at the five airports, the arrivals at RJTT, and the arrivals at RJCC in February 2009, respectively. For the overall arrivals at the five airports, the percentage of ATFM delay to pre-departure delay was around 13.5% (= 0.7/5.2). On the other hand, the percentage was 28.1% (= 1.5/5.3) for RJTT arrivals and 39.7% (= 8.5/21.4) for RJCC arrivals in February 2009. It was indicated that, for the arrivals at RJTT and RJCC in February 2009, ATFM delay accounted for a certain degree of percentage of pre-departure delay.

Bad weather conditions in which ATFM often delayed flights were outside ATFM coverage. There is a possibility that the ATFM delay for arrivals at RJCC in February 2009 could be attributed to bad weather conditions. In the sense, to study ATFM delay more precisely, data on bad weather conditions should be filtered out. To do this, data selection criteria needs to be established.

V. CONCLUSION

This paper presented results from an analysis of delays in air transport at major Japanese airports. Although long term analysis is required, the following observations were made in this trial study.

Firstly, delay was studied in terms of punctuality. Punctuality was measured based on the percentage of arrivals/departures no more than 15 minutes later than their scheduled time. Overall, the arrival punctuality was at 87.6% and departure punctuality was at 92.3%. At one of the airports, the extreme impact of weather conditions could have been a factor.

Delay was studied in terms of conformity with flight-plans. Flight operations were divided into phases. The monthly averages and the monthly standard deviation of pre-departure delay demonstrated rather wide fluctuations. Although the percentage of departures no more than 15 minutes later than the scheduled time was small, the spread of the distribution was wide. As a result, pre-departure delay proved to be the main driver of delay fluctuation. A previous study of

air transport in Europe and the United States indicated the same results[6]. The fluctuation of pre-departure delay can be attributed to the factors outside ATM coverage such as carrier-action and reactionary-delay.

ATFM delay was examined for the study of the causes of pre-departure delay. Overall, ATFM had no impact on pre-departure fluctuation. However, ATFM delay accounted for some 40% of pre-departure delay in an extreme case. At the same time, for arrivals at a domestic hub airport, ATFM delays always accounted for some 28% of delays. It should be noted that all the data were analyzed in this study regardless of weather conditions.

To study delay more precisely, the calculation method for taxi-in delay must be refined. In addition, to investigating and classifying the causes of pre-departure delay, data from airlines are required.

Delay factors, without doubt, must be continuously monitored and trends must be studied to achieve a detailed ATM performance analysis. Continuous application of the data analysis presented in this paper can assist in monitoring and controlling the delay transitions and consequently offer significant insights into future ATM improvements.

ACKNOWLEDGMENT

The authors would like to thank the Japan Civil Aviation Bureau (JCAB) for their cooperation in obtaining the analyzed data set. In particular, many thanks go to Mr. Kimihiko Itoh, Mr. Ryo Yamauchi and Mr. Akira Kimura for their cooperation.

The authors are also grateful to Dr. Kazuo Yamamoto for his useful comments.

REFERENCES

- [1] ICAO, "Global air traffic management operational concept," Doc 9854 AN/458, 2005.
- [2] J. J. Schotle, H. A. P. Blom, H. v. d. Bos, R. B. H. J. Jansen, "Management of ATM performance in operational concept development and validation : A case study," in *8th USA/Europe Air Traffic Management Research and Development*, 2009.
- [3] Performance Review Commission, "Performance review report 2006," 2007.
- [4] SESAR Consortium, "Milestone deliverable D2: The performance target," 2006.
- [5] Central Office for Delay Analysis, "Digest - annual 2008 delays to air transport in Europe," Eurocontrol, 2009.
- [6] J. Gulding, D. Knorr, M. Rose, J. Bonn, P. Enaud, H. Hegendoerfer, "US/Europe comparison of ATM-related operational performance," in *8th USA/Europe Air Traffic Management Research and Development*, 2009.
- [7] E. R. Mueller and G. B. Chatterji, "Analysis of aircraft arrival and departure delay characteristics," in *Aircraft Technology, Integration and Operations*. NASA Ames Research Center, 2002.
- [8] P. Balakrishna, R. Ganeshan, L. Sherry, "Accuracy of reinforcement learning algorithm for predicting aircraft taxi-out times," in *3rd International Conference on Research in Air Transportation*, 2008.

Analysis Of “Tarmac Delays” at Philadelphia Airport

Maricel Medina (MSc. Candidate), Lance Sherry (PhD.)

Center for Air Transportation System Research

George Mason University

Virginia, USA

mmedinam@gmu.edu, lsherry@gmu.edu

Abstract— Several incidents about passengers sitting on the runways have focused attention on Tarmac Delays. Several attempts have been done since 1990 to protect the rights of passengers. It was until December 2009 when, the Department of Transportation (DOT) approved the bill with regards to Tarmac Delays. The new regulation gives responsibilities to the airlines when passengers stay in the airplane for more than two hours. When this happens, airlines must provide food, water and lavatory service to passengers. After three hours, passengers should be allowed to return to the gate and de-plane. The right to deplane shall be waived if the pilot of such aircraft reasonably determines that the aircraft will depart or be unloaded at the terminal not later than 30 minutes after the 3 hour delay; or the pilot of such aircraft reasonably determines that permitting a passenger to deplane would jeopardize passenger safety or security. In order to investigate the occurrences of tarmac delays and effect in passengers, this paper describes the results of an analysis of “tarmac delays” at Philadelphia airport from 2005 to 2009: (a) the probability of a flight experiencing a tarmac delay of greater than 2 hours is 0.44%, (b) the average tarmac delay was 157 minutes per flight, (c) the number of tarmac delays has remained the same over the 5 year period, (d) June and July are the worst month for tarmac delays, (e) flights bound for Chicago O’Hare are the most likely to experience “tarmac delays,” (f) an estimate of the annual cost to the airlines as a result of tarmac delay regulations at PHL is \$17,000 per year or \$37 per flight.

Keywords: *Tarmac delays, apron delays, lengthy aboard aircraft waiting times, ground delay, departure delays, on-board flight delays.*

I. INTRODUCTION

Tarmac Delays is the term used to refer to flights that are “holding passengers on flight on the ground before taking off or after landing with no opportunity for its passengers to deplane”[1]. Several incidents where passengers have been stranded on airplanes for long hours have been widely publicized [2][3]. After several attempts to approve a passenger’s bill of rights, the Department of Transportation (DOT) has issued a new rule designed to protect airline passengers [4]. The 1st Session of the 111th Congress amended title 49, United States Code, to ensure air passengers have access to necessary services while on a grounded air carrier, and for other purposes [1]. The rule identifies responsibilities for three stakeholders: air carriers, airport authorities and the Department of Transportation.

The Tarmac Delay regulations require the airlines adopt and publish contingency plans for lengthy tarmac delays

including food and water for Tarmac Delays greater than 2 hours, and provides the passengers the rights to de-plane after 3 hours (with some limitations). The airlines must also respond to consumer problems, and publish tarmac delay data, designate an employee to monitor the effects of flight delays and cancellations. The strongest clause in the regulation declares “the operation of flights that remain chronically delayed to be an unfair and deceptive practice and an unfair method of competition.” The airport authority should provide a proposed contingency plan under that contains a description of: how the airport operator will provide for the deplanement of passengers following a long tarmac delay. And, the DOT shall review the initial contingency plans submitted and approve plans that closely adhere to the standards not later than six months after the date of enactment of the section.

This paper describes the results of an analysis Bureau of Transportation Statistics (BTS) [5] statistics to assess the frequency and severity of tarmac delays, and to estimate the cost of the regulations to the airlines. This study focuses on tarmac delays at Philadelphia (PHL) International Airport. Using data from 2005 to 2009, the following main results were identified:

- a) the probability of a flight experiencing a tarmac delay greater than 2 hours at PHL is 0.44%, and greater than 3 hours 0.01%
- b) the average delay experienced by the passengers was 157 minutes with a maximum time of 393 minutes (more than 6 hours),
- c) the number of tarmac delays has remained the same over the 5 year period with an average of 463 flights per year,
- d) June and July are the worst month for tarmac delays, with 49% of tarmac delays
- e) flights bound for Chicago O’Hare (14%) are the most likely to experience “tarmac delays,”
- f) an estimate of the total cost to the airlines as a result of tarmac delay regulations at PHL is \$34,000 per year or \$37 per flight.

Although the occurrence of these events is very low the severity of the delay can be high. The likelihood of tarmac delays is exogenous to the departure airport as illuminated by the most congested schedule periods (June and July) and flights

departing to the most congested airports (Chicago O'Hare - 14%, ATL - 5%, BOS - 5%). Further, the costs to the airlines are approximately only \$17,000 per year.

This paper is organized as follows: Section 2 provides a background of tarmac delays regulations. Section 3 describes the method for analysis. Section 4 provides the analysis of the tarmac delays at PHL and the implications of the rule on this airport. Section 5 summarizes the results and conclusions.

II. TARMAC DELAYS REGULATION

Tarmac delays, also known as ground delays, refer to delays that occur on the ground of the airport with passengers already on board. Tarmac delays include taxi-in, for arrivals, or taxi-out, for departures, and apron-gate (both) delays.

Regardless the cause of the ground delay, passengers face a lot of problems when a flight is held for long hours on the tarmac. To protect passengers, the DOT issued and Advance Notice of Proposed Rulemaking (ANPRM) in 2007 announcing the needs to consider or amend rules to address primarily the problems passengers face during long tarmac delays. In December 2009 that the DOT issued the regulation entitled "Enhancing Airline Passenger Protections". This regulation will take effect on April 2010 [6].

The new rule has five components to protect passengers and ensure that airlines provide a service that meet minimum standards:

1. Require carriers to adopt and publish contingency plans for lengthy tarmac delays
2. Require carriers to respond to consumer problems
3. Declare the operations of flights that remain chronically delayed to be an unfair and deceptive practice and an unfair method of competition
4. Require carriers to publish delay data on their websites.
5. Each carrier must adopt a customer service plan and self-audit adherence to it.

The first clause establishes the requirement for a contingency plans that includes:

a) Passengers on planes delayed on the tarmac for two hours will have access to food, water, clean lavatories, and the assistance of medical personal if needed, and

b) Passengers on planes delayed on the tarmac for three hours will be permitted to deplane, unless there is a safety and/or security-related impediment to deplaning passengers or air traffic control advises the pilot that permitting passengers to return to the gate or disembark would significantly disrupt airport operations. [6]

These rules apply to any carrier that operates domestic scheduled passenger service, including any charter service that uses any aircraft with 30 or more passenger seats. The contingency plan should be included for those aircrafts that

presents long ground delays on the tarmac and have fewer than 30 seats.

III. METHOD OF ANALYSIS

A data mining analysis using BTS database has been done to compute tarmac delays at PHL. The analysis includes only the use of taxi out time to compute the tarmac delays. It has been demonstrated that much of the delays occurs during taxi out because aircrafts are delayed at their origin if the predicted demand at their destination is expected to exceed the predicted capacity [7].

Fig. 1 shows the four tables and their fields used to extract the data for the analysis. All four tables were downloaded from the BTS website and stored in a local server. The local database contains information from January 2005 to October 2009. ON_TIME table keeps information about the flights. This table is the main source to compute tarmac delays by destination and day of schedule departure. ARLINE table is needed to get the name of the airline's code that exists in table ON_TIME. TAIL_AIRCRAFT_TYPE and EQUIPAGE are used to obtain information regarding the aircraft such as type, typical number of seats, and tail number.

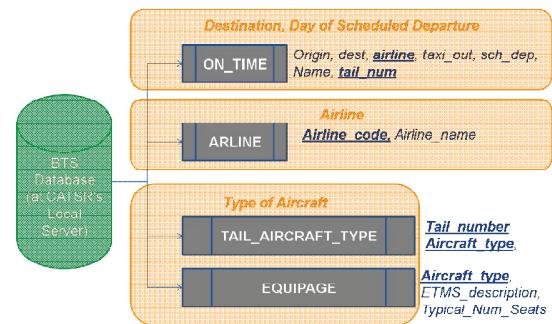


Figure 1 Data Source Model

IV. RESULTS

This section summarizes the results of the analysis.

A. Frequency of Occurrence

Table I shows the total departure flights grouped by the number of minutes delayed using taxi-out. A total of 517,887 flights departed from PHL during January 2005 and October 2009 were queried from the database. Only 0.44% of these flights, 2313, have taxi-out time greater than two hours.

The average annual percentage of delayed flights is 0.09%. The average of flights that spent more than two hours on the tarmac was 0.53% annually during 2005 and 2008. The worst year was 2007 when 0.72% of the flights spent more than two hours. Based on ten months of data for 2009, 0.70% of flights have been delayed for more than two hours on the tarmac.

TABLE I. PHL DEPARTURE FLIGHTS (JAN 2005 - OCTOBER 2009)

Tarmac Delay Grouped by Minutes								
Years	0-59m	60-119m	120-179m	180-239m	240-299m	300-359m	360-419m	TOTAL
2005	122,278	3,620	456	69	6			126,429
2006	105,045	2,833	339	85	10			108,312
2007	99,729	3,729	445	122	31	4		104,060
2008	97,372	2,770	248	81	27		1	100,499
2009	75,554	2,644	329	54	6			78,587
TOTAL	499,978	15,596	1,817	411	80	4	1	517,887
	96.54%	3.01%	0.35%	0.08%	0.02%	0.00%	0.00%	100%
Total Flights with Tarmac Delay >= 2 hours								2,313
Percentage of Flights with Tarmac Delay >= 2 hours								0.44%

Only five flights of the total delayed flights (0.17%) have spent more than four hours on the tarmac during five years of analysis. This represents only 0.0007% of total flights (Fig. 2). During the analyzed years, 96.54% of the flights stayed less than one hour on the tarmac, only 3.01% stayed between one and two hours, 0.35% between two and three hours and only 0.10% three hours or more (Fig. 2).

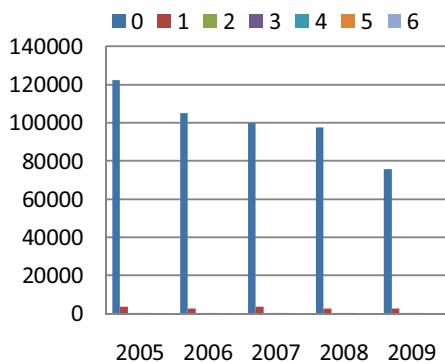


Figure 2. Total Flights with a Tarmac delay (no cumulative)

C. Tarmac Delays by Airline

Fig. 3 shows the relation between airlines and tarmac delays. 1123 out of 2313 flights (49%) were operated by US Airways, followed by Southwest Airlines with 15%. This may be happening because PHL is the primary hub for international flights of US Airways. Also, the only flight that stayed for more than 6 hours on the tarmac in 2008 corresponds to the US Airways flight US598, an A320 aircraft with destination to LAS. Its taxi-out time was 393 minutes reported as reported on the BTS, and was scheduled to departure on Monday, July 23rd at 4:05:00 PM.

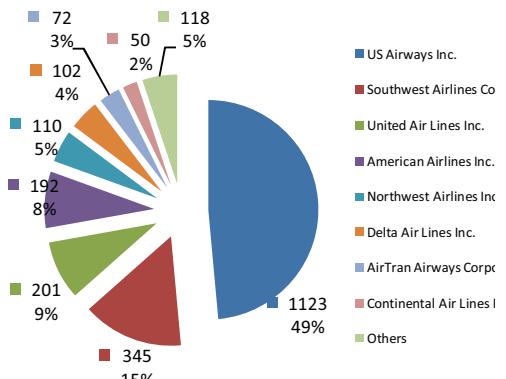


Figure 3. Tarmac Delays by Airlines

B. Severity of Delays

The average delay is 157 minutes (2.37 hours), the minimum delay is 120 minutes, the maximum is 393, and the mode is 120 minutes. Figure 4 Fig. 4 shows the distribution of flights and the average taxi-out time for each tarmac delay group at PHL airports during the analyzed period.

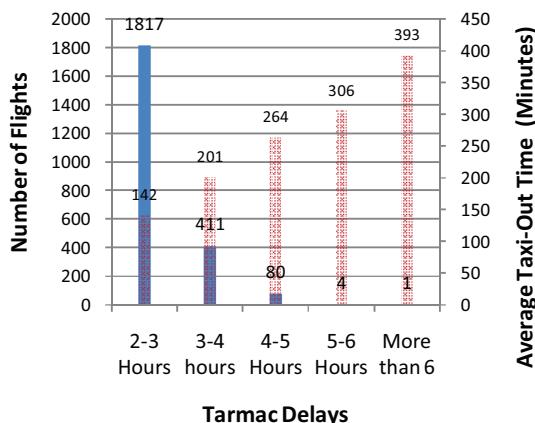


Figure 4. Distribution of Tarmac Delays and Average Taxi-Out Time

Only four flights, two from US Airways and two from Southwestern Airlines, had taxi-out times greater than five hours (Table II). Three out of these flights occurred the same day, on Saturday February 14th, 2007 during the morning. That day a severe snow storm was affecting the operations at the airport.

TABLE II. TAXI-OUT TIMES GREATER THAN 5 HOURS (JANUARY 2005, OCTOBER 2009)

Flight #	Destination	Schedule Departure	Aircraft Seats
US1071	SJU – San Juan, Puerto Rico	02/14/2007 9:30:00 AM	A333 - 295
US1991	ORD - O'Hare Int'l Airport	02/14/2007 9:50:00 AM	B733 - 128
WN993	MCO – Orlando Int'l Airport, FL	02/14/2007 7:10:00 AM	B737 – 126
WN2276	MDW - Chicago	6/19/2007 5:15:00 PM	B737 – 126

D. Tarmac Delays by Destination

Flights with destination to ORD have had the greater taxi-out times. Table III shows the four airlines that flew to ORD and its taxi-out time has been greater than two hours. This represents only a 13.57% of the delayed flights and 0.06% of total flights during the analyzed period. In this case, American Airlines' flights have been stayed more on the ground than US Airways flights.

TABLE II. AIRLINES WITH TARMAC GREATER THAN TWO HOURS AND ORD AS DESTINATION (JANUARY 2005- OCTOBER 2009)

Year	AA-American Airlines	MQ-American Eagle	UA-United Airlines	US-US Airways	Grand Total
2005	23		28	17	68
2006	26	1	29	19	75
2007	32		28	19	79
2008	10		18	22	50
2009	6	5	16	15	42
Total	97	6	119	92	314
%	30.89%	1.91%	37.90%	20.30%	100%

E. Tarmac Delays by Month

The analysis shows that most of the tarmac delays flights happen during June and July. It seems to be a strong relationship between summer and taxi-out time (Fig. 5).

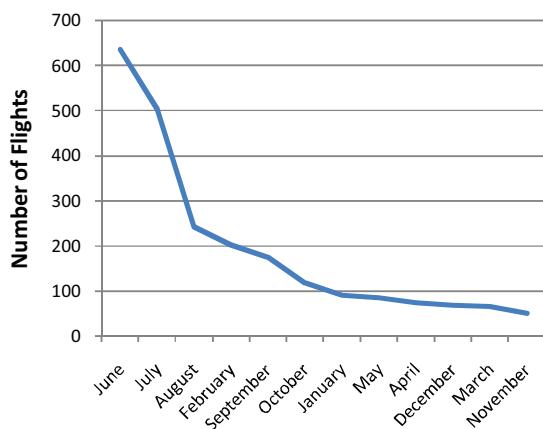


Figure 5. Total Number of Flights By Month with Tarmac greater than two hours

F. Monetary Cost Analysis based on Component One

Four cost categories have been defined to estimate how much it would cost to the airlines to meet the above requirements (Table III). The costs only include labor costs and not any additional costs such fuel. Note that the regulation requires airlines to provide:

- provide food, water and medical assistance when the tarmac delay exceeds two hour, and
- allow passengers to deplane after three hours.

TABLE III. COST CATEGORY AND PRICE PER PASSENGER

Cost Category	Price Per Passenger
Food and Drinks	\$3.71
Deplaning	\$1.37
Re-boarding	\$1.37
Airfare average cost at PHL	\$220

Table IV shows the number of passengers per tarmac delay. To compute the costs to airlines on each category, the following formula is used:

$$\text{Cost Category} = \text{Total Passengers using load Factor 80\%} * \text{cost's fee}$$

TABLE IV. ESTIMATED NUMBER OF PASSENGERS AND TOTAL PASSENGERS ON TARMAC DELAYS

Tarmac Delay Criteria	Number of Passengers (January 2005- October 2009)*	Total Passengers using Load Factor 80%
2 Hours	252,077	172,795
3 Hours	71,315	57,052
4 Hours	11,970	9,576
5 Hours	675	540
6 Hours	150	120
Total passengers		268,950
Average passengers per flight		116

NOTE: Total number of passengers is computed by multiplying number of flights by number of average seats of the aircraft. When BTS does not include the aircraft type, 145 is used as the average number of seats.

Table V summarizes the costs per category based on the number of passengers and tarmac delay. The total annual average cost per airline is \$17,711.

TABLE V. COST PER CATEGORY TO MEET COMPONENT ONE

	Food	Deplane	Re-Board	Cancellation
Passengers	268,949.60	67,288.00	65,403.94	1,884.06
Cost per passenger	\$3.71	\$1.37	\$1.37	\$220.00
Total Cost (Jan 2005- October 2009)	\$997,803	\$92,184	\$89,603	\$414,493
Annual Cost	\$199,560	\$18,437	\$17,920	\$82,898
Average per airline (18 airlines)	\$11,087	\$1,024	\$995	\$4,605
Total Average Annual Cost				\$17,711

NOTE: Subtotals may not add due to rounding

- Food Costs: Food costs were computed based on the total number of passengers with a tarmac delay greater than two hours. The estimated annual average cost per airline is \$11,087.
- De-plane Costs: De-plane costs only occur when flights have a taxi-out greater than three hours. The total number of deplaned passengers is 67,288, representing an average cost of \$1,024 per airline.
- Re-Board Costs: This cost is computed based on the probability of re-boarding an airplane which is 97.2% [8]. The total number of re-boarding passenger is 65,403 with an average cost of \$995 per airline.
- Cancellation: This cost is computed based on the probability of cancelling a flight after deplaning which is 2.8% [8]. The total number of passengers in flights cancelled after deplaning is 1,884. The average airfare cost at PHL is \$220 [5]. The average annual cost for an airline to cancel flights when taxi-out time is greater than three hours is \$4,605.

V. CONCLUSIONS

In the last decade there have been several widely publicized incidents in which passengers were on flights for extended periods with degraded comfort including the absence of food and service amenities, and were unable to deplane [9].

Although the occurrence of these events is very low the severity of the delay can be high. Tarmac delays in excess of 2 hours occur 0.44% of the time. Tarmac delays in excess of 3 hours occur 0.01%. The average delay was 157 minutes. The most congested periods are during summer (June) and to the most congested airports (Chicago O'Hare -14%, ATL – 5%, BOS – 5%).

Further, the cost to the airlines is approximately \$17,711 annually. A similar analysis has been done for New York airports [9]. Further analysis will include OEP35 airports.

ACKNOWLEDGMENT

Thank you for technical assistance and suggestions from John Ferguson, Guillermo Calderón (CATSR). This project was funded by internal George Mason University Foundation Funds.

REFERENCES

- [1] Camara of Representatives, Airline Passenger Bill of Rights Act of 2009, 111TH Congress, 1st Session: 2009.
- [2] "FlyersRights.ORG - Largest Non-Profit Airline Consumer Organization. <http://flyersrights.org/>."
- [3] "Business Travel Coalition (BTC) Home Page. <http://businesstravelcoalition.com/>."
- [4] B. Goldberg and D. Chesser, Sitting on the Runway: Current Aircraft Taxi Times Now Exceed Pre-9/11 Experience, Bureau of Transportation Statistics, 2008.
- [5] "RITA | Bureau of Transportation Statistics (BTS). <http://www.bts.gov/>."
- [6] Department of Transportation, Enhancing Airline Passenger Protections, Federal Register, 2009.
- [7] Evans, Antony David, "Responses to Airport Delays -A system study of Newark International Airport," Massachusetts Institute of Technology, 2002.
- [8] Decision Economics, Final Regulatory Impact Analysis of Rulemaking on Enhanced Airline Passenger Protections, 2009.
- [9] M. Medina-Mora and L. Sherry, "Analysis of Tarmac Delays at New York Airports," Integrated Communications Navigation and Surveillance Conference (ICNS), Herndon, VA, USA: 2010.

Trajectory Prediction by Functional Regression in Sobolev Space

K. Tastambekov*, S. Puechmorel*, D. Delahaye* and C. Rabut[†]

*Ecole Nationale de L'Aviation Civile, Dept. Mathematiques et Informatique LMA,

7, Avenue Edouard Belin, 31055 Toulouse France

[†] Institut National Des Sciences Appliquees de Toulouse, Dept. Mathematiques,

135, Avenue de Rangueil - 31077 Toulouse Cedex 4 France,

email: christophe.rabut@insa-toulouse.fr

Abstract—In this paper we consider the problem of short to mid-term aircraft trajectory prediction. That problem aims to predict collisions between aircraft in airspace. Our approach is based on local functional regression which consists in the three following stages : data pre-processing, localizing and regression. This algorithm has been successfully applied on aircraft trajectories between Toulouse and Paris.

Index Terms—Trajectory prediction, wavelet, functional regression.

I. INTRODUCTION

A. Trajectory Prediction Metrics

Air traffic management research and development has developed substantial collection of decision support tools (DST) that provide automated conflict detection and resolution, trial planning, controller advisories for metering and sequencing, traffic load forecasting, weather impact assessment. Aircraft trajectory prediction algorithms([25]) are significant components of decision support tools (DST) in order to avoid collisions with others aircraft, arrival metering and other applications in air traffic management. A 4-dimensional (4D) trajectory prediction contains data specifying the predicted horizontal and vertical position of an aircraft over some time span into the future. The ability to accurately predict trajectories for different types of aircraft and under different flight conditions, that involves external actions (pilot, ATC) and atmospheric factors (wind, temperature), is an important factor in determining the accuracy and effectiveness of an air traffic management. Everyday, about 8000 aircrafts fly in the French airspace, inducing a huge amount of control workload (see [27]). Such workload, is then spread by the mean of the airspace sectoring. The airspace is divided into geometrical sectors, each of them being assigned to a controller team. When a conflict between two (or more) aircraft is detected, the controller changes their routes (heading, speed, altitude) in order to keep a minimum distance between them during

the crossing. All flying aircrafts are then monitored during their navigation and so from the departure till the destination. When a controller observes its traffic on the radar screen, he tries to identify convergent aircraft which may be in conflict in a near future, in order to apply maneuvers that will separate them. The problem is to estimate where the aircraft will be located in a near future (10 – 30 minutes).

One of the issues in trajectory prediction is to measure how accurately a model will fit to a target trajectory. Unfortunately, many different metrics can be proposed each of them focusing on a specific aspect of accuracy. Most of the time, the proposed metrics fall into one of these categories [26], [27]:

- Time coincidence. The time difference between a predicted event and a real event is used as a measure of TP accuracy. Time coincidence is relevant in applications where synchronizing is important, like sequencing traffic, or when the DST uses time information to inform controller about the actions that have to be taken.
- Spatial coincidence. Similar to the previous one except that spatial distance at specified time (or more generally at events that can be predicted with the knowledge of aircraft positions up to a given time) between the model and the real aircraft is computed. Spatial coincidence can be refined by further splitting into altitude and horizontal error. Furthermore, for some applications, mainly conflict predictors and/or solvers, spatial difference is projected onto a vector normal to the real trajectory (cross-track error) and onto a vector tangent to the real trajectory (along-track error).
- 4D coincidence. Trajectories are considered as 4D curves, and distance between such curves is computed. Most of the metrics derived for spatial coincidence can be extended to the 4D setting, with the benefit of including a kind of time coincidence, thus generalizing in some sense the previous two aspects.
- Morphological similarity. Different in nature from the previous metrics, an intrinsic distance between trajectories considered as curves in a 3D space can be derived from Riemannian geometry. Since only the shape of the trajectory is taken into account, this metric is relevant

mainly for trajectory design tools.

Except for the last one, all those basic metrics can be integrated along trajectories to produce a mean value indicator (the classical L^2 distance is for example obtained by integrating the standard spatial coincidence metric over time).

B. What is Functional Data Analysis?

Functional data analysis is an active branch of statistics in which relevant objects are mappings belonging to a well defined space, most of the time a Hilbert space. It has been proved very efficient for problems where preserving the functional nature of data is of great importance: curves classification, functional dependence learning and similar problems. The fundamental aims of functional data analysis are the same as those of conventional statistics ([12], [13], [14], [15], [16]):

- to formulate the problem at hand in a way amenable to statistical thinking and analysis
- to develop ways of presenting the data that highlight interesting and important features,
- to investigate the variability as well as mean characteristics,
- to build a model for observed data , including those that allow for dependence of one observation or variable of another, etc.

In some cases, original observations are interpolated longitudinal data which are quantities observed as they evolve through time. In other situations, people prefer to use panel data, which are data from a number of observations over time of cross-sectional units like individuals, households, in our case aircraft trajectories.

We encounter functional data in many applications. In our problem, data consist of large number of independent numerical observations coming from ATC radars. Such data represent aircraft trajectories in the French airspace. In recent papers an increasing attention has been paid to linear functional regression, and some of its generalizations. In this setting, either a scalar value or a mapping (the response), possibly contaminated by an independent measure noise is modeled as being linearly dependent on a mapping (the predictor).

In this paper we will present an innovative approach based on functional regression for solving short to mid-term trajectory prediction (TP) problem. The first part of the paper presents our approach and give the associated mathematical modeling. The second part presents initial results on real data for the flight from Toulouse to Paris.

II. LOCAL FUNCTIONAL LINEAR REGRESSION BASED ON

WEIGHTED DISTANCE-BASED REGRESSION

A. Problem Statement

The main idea of this paper is to solve the linear functional regression problem using data coming from radar tracker in order to build an enhanced trajectory prediction. An aircraft

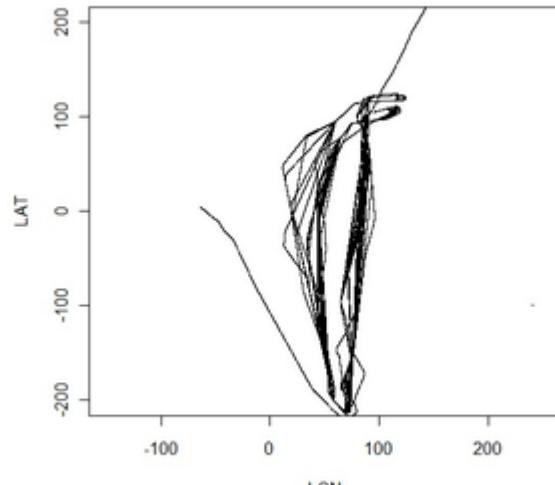


Figure 1. Example of aircraft trajectories data (Toulouse-Paris).

trajectory is by definition a mapping from a time interval $[a, b]$ to \mathbb{R}^3 .

Radar data are disturbed by noise measurement and are not regularly sampled in time. These data are then processed by smoothing, approximation, and resampling. This is the first step of the algorithm.

A flight path is controlled by flight dynamics equations and knowing that the pilot's actions are simple, we may assume that aircraft trajectories are C^1 , piecewise C^2 functions. Thus, we may assume that observed trajectories are samples of an Hilbert stochastic process (in fact it is even a Sobolev space valued process).

Let $\{X_n, Y_n\}_{n=1}^N$ be a sample of observations identically distributed coming from a Hilbert random processes X, Y defined on intervals τ_X, τ_Y . For our application, X_n represents all past trajectories connecting the same origin destination and Y_n the associated "future" reference trajectories. τ_X is the past time horizon on which we gather trajectory samples and τ_Y is the time horizon on which we do the prediction.

In order adjust our model, real trajectories will be used (from Toulouse to Paris; an example of radar tracks picture of such trajectories can be found on figure 1) for which a reference time will be considered. This time will artificially separate the past and the "future" (which is known in this framework). Based on the previous position ($\{X_n\}$), our model will produce future predicted positions $\hat{Y}(t)$ in order to minimize the errors between Y_n and $\hat{Y}(t_n)$ see figure 2.

Let μ_X, μ_Y and B_X, B_Y be means and covariance kernels respectively (μ_X represents the mean trajectory on the set $\{X_n\}$ on τ_X and B_X the associated standard deviation).

The functional linear model has the general form [26]:

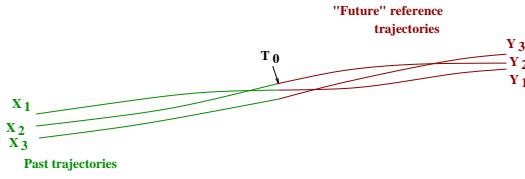


Figure 2. Framework used to adjust the model. Our model is trained by the mean of a set of trajectories executed between two airport (in our case Toulouse-Paris). A reference time T_0 is then considered to separate the past and the future. The model uses both parts in order to adjust its parameters. After this training phase, the algorithm is tested by using only the samples coming from the past to predict the future position of the aircraft. Having the reference trajectory, it is very easy to measure the accuracy of the prediction.

$$\hat{Y}(t) = \hat{f}(t) + \int_{\tau_X} \hat{K}(t, s) X(s) ds$$

where $\hat{f}(t)$ a smooth square integrable mapping which represents the mean of learning trajectories data set and $\hat{K}(t, s)$ is a smooth square integrable matrix valued kernel. Next integral:

$$\int_{\tau_X} \hat{K}(t, s) X(s) ds$$

is the deviation of the predicted trajectory from the mean $\hat{f}(t)$.

The solution of the functional regression problem is the optimal couple (\hat{f}, \hat{K}) that minimize the mean square error between Y and \hat{Y} . Several expansions has been tried to model trajectories, but after some experiments we decide to use wavelet decomposition it is the which minimizes the prediction error. In our algorithm we use such wavelet in Sobolev space instead of the regular L_2 ([19], [20], [21]).

B. Wavelets in Sobolev space

At the beginning of 1980s, many scientists were already using "wavelets" as an alternative to traditional Fourier analysis. The word "wavelet" is used in mathematics to denote a kind of orthonormal bases in L_2 with remarkable approximation properties. The theory of wavelets was developed by Y.Meyer, I.Daubechies, S.Mallat and others in the end of 1980s, [1], [2], [3], [4], [5].

Definition 1: Sobolev space. Let $s \in \mathbb{N}$. The function $f \in L_2(\mathbb{R})$ belongs to the Sobolev space $W^s(\mathbb{R})$ [17], [18], if it is s -times weakly differentiable, and if $f^{(j)} \in L_2(\mathbb{R}), j = 1, 2, \dots, s$. In a Sobolev space the norm is given by :

$$\|f\|_{W^s(\mathbb{R})}^2 = \|f\|_{L_2(\mathbb{R})}^2 + \|f^{(s)}\|_{L_2(\mathbb{R})}^2$$

Any $f \in L_2(\mathbb{R})$ can be represented as a series (convergent in $L_2(\mathbb{R})$) which is the definition of standard wavelet decomposition [6], [7], [8], [9] :

$$f(t) = \sum_{k \in \mathbb{Z}} c_k \varphi_k(t) + \sum_{j=0}^{\infty} \sum_{k \in \mathbb{Z}} c_{jk} \psi_{jk}(t) = \sum_{i \in \mathbb{Z}} c_i \varphi_i(t)$$

where c_k, c_{jk} are some coefficients, and

$$\|f\|_{L_2(\mathbb{R})}^2 = \sum_{k \in \mathbb{Z}} c_k^2 + \sum_{j=0}^{\infty} \sum_{k \in \mathbb{Z}} c_{jk}^2.$$

It was shown in [1],[4], that a function f lies in $W^s(\mathbb{R})$ if and only if

$$\sum_{k \in \mathbb{Z}} c_k^2 + \sum_{j=0}^{\infty} \sum_{k \in \mathbb{Z}} 2^{sj} c_{jk}^2 < +\infty.$$

Moreover, the discrete equivalent norm in Sobolev space $W^s(\mathbb{R})$ is

$$\|f\|_{W^s(\mathbb{R})}^2 \approx \sum_{k \in \mathbb{Z}} c_k^2 + \sum_{j=0}^{\infty} \sum_{k \in \mathbb{Z}} 2^{sj} c_{jk}^2$$

where s is the smoothness order of the Sobolev space.

For our application, trajectories will be modeled by such decomposition (wavelet in Sobolev space)[23].

III. APPLICATION TO TRAJECTORY PREDICTION

This section presents the application of the previous modeling to the trajectory prediction problem. The wavelet decomposition has been extended to the 3-dimensional case in order to fit the aircraft trajectories. Several kind of wavelets have been tried, but Daubechies 4 has produced more better results. The first part presents the solving of the function regression and the second one gives the results of our algorithm on real data set.

A. Solving the Functional Regression

One of the goal of our approach is to select the Hilbert random process from the set of trajectories connecting different points (airports in our application) [26]. As we said above, for French airspace only, there are about 8000 aircrafts every day connecting many Origins-Destination pairs. Without the knowledge of the origins and destination airports, trajectory prediction problem is much harder to address. It means that we have to extract a subset of trajectories connecting the same origins-destination pair. For our application, we have decided to keep the tracks of the flights from Toulouse to Paris for a given day (May 8 2009).

Let X_n (response Y_n) be the realization of predictor process X (response for process Y) corresponding to observation n in the data set. Let M_n (response L_n) be the number of samples available for this observation (number of radar plots of a given trajectory) and let $X_{n,j}, j = 1, \dots, M_n$ (response $Y_{n,j}, j = 1, \dots, L_i$) be the actual samples along trajectories X_n

(response Y_n) with corresponding sample times $\tau_{n,j}$ (response $\nu_{n,j}$). The number of samples M_n , L_n and the sampling times are assumed to be random variables independent from the processes X and Y. The first step towards solving the problem is to resample time intervals and to compute the missing data. To make an expansion of the predictor and response on respective basis and to compute coordinates at a given fixed time interval we can use several basis representations, such as wavelets, cubic splines, karhunen-loeve expansion, etc. An important step in the design of a linear smoother is the choice of weighted kernel and bandwidth. The problem has been addressed in the field of non parametric statistics and it is known that the kernel has less influence than the bandwidth. Let us now introduce some examples of kernels. The Epanechnikov kernel is defined by the following equation :

$$K_e(t) = \frac{3}{4}(1-t^2)1_{[-1,1]}(t)$$

This kernel has some interesting optimality properties and is easy to compute. Another choice is the Gaussian kernel :

$$K_g(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$$

For very fast computation, it is still possible to use a uniform kernel :

$$K_u(t) = \frac{1}{2}1_{[-1,1]}(t)$$

Since the data set is usually large (around 1000 trajectories sampled at 10s), a compactly supported kernel in the local linear smoother allows a reduced computational load and a complexity mostly independent of the number of samples of the trajectory. The Gaussian kernel is not compactly supported, but decreases very fast at infinity so that practically it can be set to 0 outside a compact interval. And as a simple particular case we can define the simple window function:

$$W_d(X_i, X_j) = \begin{cases} 1, & \text{if } d(X_i, X_j) \leq d \\ 0, & \text{if } d(X_i, X_j) > d \end{cases}$$

Where $d(X_i, X_j) = \|X_i - X_j\|$ is the distance between two trajectories, X_i and X_j respectively.

Finding the right predictor is a critical task in applying functional regression. For trajectory prediction purpose, it is natural to consider a part of the observed trajectory as the learning set, and a part of the future trajectory as target. The learning database has thus been chosen by selecting homogeneous 256 radar plots (per trajectory) from a day of traffic. Those learning trajectories have been divided into two parts with 128 plots each, corresponding to the past and the “future” .

One simple approach to estimate $f(t)$ is to center the observed Y_n and the given X_n by subtracting their sample average functions \bar{X} and \bar{Y} . Here and later we consider centralized X_n, Y_n and the model becomes :

$$\hat{Y}(t) = \int_{\tau_x} \hat{K}(s, t)X(s)ds$$

Then, the regression problem becomes to find an optimal $\hat{K}(t, s)$ minimizing following expression:

$$\sum_{n=1}^N U_d \|Y_n(t) - \int K(t, s)X_n(s)ds\|_{(W^1)^3}^2$$

where U_d is one of the weighted window kernel function described above([24]).

The kernel $\hat{K}(t, s)$, $X_k(s)$ and $Y_k(t)$ can be expressed using the wavelet basis $(\phi_i)_{i \in \mathbb{N}}, (\psi_i)_{i \in \mathbb{N}}$ as:

$$X_n(s) = \sum_j a_j^n \phi_j(s), \quad Y_n(t) = \sum_i b_i^n \psi_i(t)$$

$$K(t, s) = \sum_i \sum_j K_{ij} \phi_j(s) \psi_i(t)$$

Where ϕ_i and ψ_j are wavelet basis functions, respectively to the τ_X and τ_Y time intervals. Using the orthonormality of basis, the regression problem becomes to find the minimum of the sum:

$$\min_{K_{ij}} \sum_{n=1}^N U_d \|Y_n(t) - \int K(t, s)X_n(s)ds\|_{(W^1)^3}^2 =$$

$$\min_{K_{ij}} \sum_{n=1}^N U_d \sum_{i=1}^P \sum_{j=1}^Q (b_{in} - \sum_{j=1}^Q a_{jn} K_{ij})^2$$

Here the expansions were truncated to a fixed rank. Then $\hat{f}(t)$ can be founded by the next formula:

$$\hat{f}(t) = \bar{Y}(t) - \int_{\tau_X} \hat{K}(t, s) \bar{X}(s) ds$$

$$= \sum_{i=1}^P (\bar{b}_i - \sum_{j=1}^Q \hat{K}_{ij} \bar{a}_j) \psi_i(t)$$

which is nothing but a linear mean square problem that can be solved with the help of normal equations or using SVD.

B. Application to real data

For the first test we use one day air traffic between Toulouse and Paris airports. There are 52 aircraft trajectories flying in both directions and one trajectory will be used as “real” trajectory for which prediction accuracy will be evaluated. Each of the 52 trajectory will be selected as “real” to build a cross validation procedure in order to improve robustness. The least mean square problem was solved using window kernel function. Figure 3 shows the first 42 minutes of real aircraft trajectories started from Toulouse-Blagnac airport. And the second figure 4 consists of two parts. The first part shows 21 minutes of real trajectories and the second part is the predicted trajectories. The shift is located at the middle of trajectories (50 on the Y axis).

A cross validation procedure was designed as follows:

- Pick a trajectory and remove it from the leaning database.

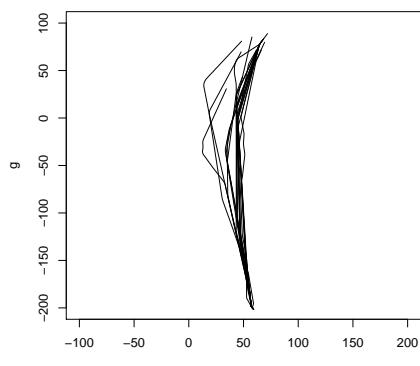


Figure 3. Real trajectories

Relative errors			
0.1622	0.2269	0.2020	0.1848
0.1719	0.1994	0.1935	0.1940
0.2176	0.1653	0.1812	0.1636
0.1940	0.1553	0.2977	0.2090
0.2168	0.1686	0.1691	0.9003
0.2214	0.2393	0.3565	0.2039
0.2020	0.1681	0.1498	0.1841
0.3023	0.2183	0.1734	0.2410
0.1679	0.1622	0.2502	0.2880
0.1620	0.1847	0.2920	0.1911
0.2901	0.1968	0.1960	0.1788
0.2345	0.1969	0.1861	0.2205
0.2458	0.2873	0.1636	0.1767

As we can see, the results produced by this new approach are very good for a prediction horizon of 20 minutes.

IV. CONCLUSION AND FUTURE WORK

The functional data analysis has been applied in order to build a new algorithm for aircraft trajectory prediction problem. This approach uses only previous radar tracks for a given origins destination pair. A learning process enable the adjustment of parameters. This model is based on localization of functional linear regression model using wavelets in Sobolev space. This method produces efficient results with hight robustness.

In a next step, a larger data base will be used for the same origins destination pair (Toulouse-Paris), by taking several years of data. Then, we will try it on some other origins destination pair (midge range and long range). Finally, we will determine the limit of this approach by increasing the prediction time interval.

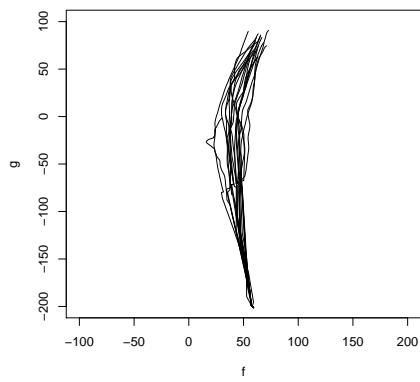


Figure 4. Predicted trajectories

- Compute prediction error on this trajectory using the others as learning set.
- Do the same with couples of trajectories removed.
- The relative prediction error computed for each trajectory is given by:

$$\frac{\|\hat{X}_0 - X_0\|^2}{\frac{1}{N} \sum_{n=1}^N \|X_i - \bar{X}\|^2}$$

where X_0 is a real trajectory, \hat{X}_0 is predicted and $\{X_n\}_1^N$ is the set of learning trajectories.

Results are summarized in the following table for which each cell give the prediction relative error for a given “real” trajectory :

REFERENCES

- [1] I. Daubechies, Ten lecture on wavelets, *SIAM*, 1992.
- [2] I. Daubechies, The wavelet transform, time-frequency localization and signal analysis, *IEEE Trans. Inf. Th.*, 1990.
- [3] C.K. Chui, An Introduction to wavelets, *Xian Jiaotong Univ. Press*, 1994.
- [4] Y. Meyer, Wavelets and Operators, *Cambridge Univ. Press*, 1992
- [5] Y. Meyer, Wavelets: Algorithms and Applications, *SIAM*, 1993.
- [6] W. Hardle, D. Picard, G. Kerkyacharian and A. Tsybakov, Wavelets, Approximation and Statistical Applications, *Un premier resultat du seminaire Paris-Berlin*, 1997.
- [7] S.A. Dianat, R. Rao, Wavelet transforms: Theory and applications, *Proc. SPIE 12th Annu. Int. Symp. Aerosense*, 1998.
- [8] M.A. Cody, The fast wavelet transform, *Dr. Dobb's J.*, 1992.
- [9] Rong-Qing Jia, Jianzhong Wang and Ding-Xuan Zhou, Compactly supported wavelet bases for Sobolev spaces, *Appl. Comput. Harmon. Anal.* 15, 2003.

- [10] I.T. Jolliffe, Principal component analysis, *Springer-Verlag*, 1986.
- [11] J.E. Jackson, A user's guide to principal components, *John Wiley*, 1991.
- [12] J. Ramsay, B. Silverman, Functional Data Analysis, *Springer series in statistics*, 1997.
- [13] J. Ramsay, B. Silverman, Functional Data Analysis (second edition), *Springer series in statistics*, 2005.
- [14] F. Ferraty, Ph. Vieu, Nonparametric functional data analysis, *Springer series in statistics*, 2006.
- [15] J. Ramsay, G. Hooker, S. Graves, Functional data analysis with R and MATLAB, *Springer series in statistics*, 2009.
- [16] J. Ramsay, B. Silverman, Applied functional data analysis: Methods and case studies, *Springer series in statistics*, 2002.
- [17] R. Adams, Sobolev spaces, *Academic press*, 1975.
- [18] I. Gihman and A. Skorohod, Introduction to the theory of random processes, "Nauka", 1965
- [19] Maya R. Gupta, Nathaniel P. Jacobson, Wavelet principal component analysis and its application to hyperspectral images, *IEEE*, 2007.
- [20] Feng Lu, Zhaoxia Yang and Yuesheng Li, Wavelets approach in choosing adaptive regularization parameter, *Lecture Notes in Computer Science*, 2007.
- [21] R. A. Devore, Fast wavelet techniques for near-optimal image processing, *IEEE Military Communications Conference Record*, 1992.
- [22] Bhavik R. Bakshi, Multiscale PCA with application to multivariate statistical process monitoring, *AIChE Journal*, 1998.
- [23] R. Suyundykov, S. Puechmorel, L. Ferre, Multivariate Functional Data Classification using wavelets in Sobolev space, *Ecole Nationale de l'Aviation Civile*, 2010
- [24] Eva Boj, Pedro Delicado, Josep Fortiana, Local Linear Functional Regression based on Weighted Distance-Based Regression, *Universitat de Barselona, Universitat Politècnica de Catalunya*, 2008
- [25] Chester Gong, Dave McNally, A methodology for automated trajectory prediction analysis, *San Jose State University, NASA Ames Research Center, Moffett Field, CA, 94035, NASA Ames Research Center, Moffett Field, CA, 94035*, 2004
- [26] D. Delahaye, S. Puechmorel, L. Boussouf, Trajectory prediction: the functional regression approach, *Toulouse, France*
- [27] S. Mondoloni, S. Swiertstra and M. Paglione, Assessing trajectory prediction performance metrics definition, 24-th Digital and Avionics Systems Conference, 2005.

A New Method for Generating Optimal Conflict Free 4D Trajectory

Nour Dougui
 Applied Mathematics Laboratory
 ENAC
 7, Avenue Edouard Belin
 31055 Toulouse, France
 Email: nour@recherche.enac.fr

Daniel Delahaye
 and Stephane Puechmorel
 Applied Mathematics Laboratory
 ENAC
 7, Avenue Edouard Belin
 31055 Toulouse, France
 Email: delahaye@recherche.enac.fr
 Email: puechmor@recherche.enac.fr

Marcel Mongeau
 Université de Toulouse UPS, INSA, UT1, UTM
 Institut de Mathématiques de Toulouse
 F-31062 Toulouse cedex 9, France
 CNRS
 Institut de Mathématiques de Toulouse UMR 5219
 F-31062 Toulouse cedex 9, France
 Email: mongeau@math.univ-toulouse.fr

Abstract—The need for increasing air traffic capacity motivates 4D trajectory planning concept. In order to generate conflict-free 4D trajectories, we introduce a new concept based on light propagation modeling algorithm. This algorithm is a wavefront propagation method that yields a natural solution for the path planning problem specifically in the case of air traffic congestion. We conclude this paper with numerical experimentation of our approach on a simplified (2D + time) test problem.

I. INTRODUCTION

The analysis of air traffic growth expects a doubling of the number of flights over the next 20 years. The Air Traffic Management (ATM) will therefore have to absorb this additional burden and to increase the airspace capacity, while ensuring at least equivalent standards of safety and interoperability. The European project SESAR was initiated to propose solutions to this problem. It relies on a new concept of air traffic control, known as 4D (3D + time) trajectory planning, which consists in exploiting the possibilities of the Flight Management System (FMS) to ensure that a given aircraft is at a given position at a given time. For each flight, a reference trajectory, called Reference Business Trajectory (RBT), is requested by the operating airline. During the flight, conflict situations may nevertheless occur, in which two or several aircraft can dangerously approach each other. In this case, it is necessary to modify one or more trajectories to ensure that minimum separation standards (currently 5 Nm horizontally and 1000 ft vertically) are still satisfied. Moreover, it is desirable that proposed new trajectories deviate as little as possible from RBT. In this context, we propose a new algorithm which seeks to ensure sufficient separation between aircraft while producing flyable trajectories.

A. Previous related works

During recent years, several methods have been proposed to find an optimal solution that could solve conflicts in air traffic. The aim of these methods is to find for each aircraft, an optimal 4D trajectory that avoids conflicts with other aircraft, reaches the destination point and optimizes a cost function which depends on the travel duration and on the cost index (a

coefficient that takes into account fuel consumption). There are mainly two classes of methods to address this problem: genetic algorithms [1] and navigation-function based approach [2]. Each one provides only a partial solution to the problem.

The first one, genetic algorithms, consists in generating a new population of aircraft trajectories from a base population using three basic operators: selection, mutation and crossover in order to improve the cost function. This process is iterated until the cost function is no longer improved. The state space is a set of finite maneuvers, which are straight lines, turning points (changing an aircraft heading and then bringing it back on its initial trajectory) and offsets (inducing a lateral shift from the initial trajectory). Those maneuvers are the ones used by air traffic controllers. Genetic algorithms generate trajectories with feasible operational maneuvers and with velocities within bounded ranges. They can reach asymptotically optimal solutions, but for a given computing time, a feasible (conflict-free) solution is not guaranteed.

The second method, based on navigation functions, consists in using an electrostatic modeling of the problem: an electron (which has a negative charge) is subject to an electric field, and is attracted by a positive charge which represents the goal and is pushed away by negative charges which represent obstacles. Thus, the electron is going to move towards the goal and steered by the resultant electric fields. The aircraft (a virtual electron) is represented by a point in 3D space. If a mathematical function of potential fields can be built to model adequately the destination charge and the distribution of the obstacle charges, then the virtual forces applied on such virtual electron, initially positioned at the departure point, can be computed. This produces a trajectory which connects the departure point with the destination while avoiding obstacles (the other aircraft).

Navigation functions have already demonstrated their effectiveness in motion planning with guaranteed collision avoidance and convergence towards the goal configuration (to reach the destination point with the right orientation). However, they do not take into account the constraints imposed by ATM, such as bounded speed, smooth trajectory and time constraints.

Besides, they may tolerate large deviations from RBT.

The objective of our approach, based on an optical analogy, is to find for each aircraft a feasible (relevant to ATM constraints) optimal 4D trajectory, avoiding conflicts and which minimizes a criterion based on a local metric.

B. Paper overview

In the next section, we present our method. Numerical results are presented in section III. Then, a conclusion appears in section IV.

II. LIGHT MODELLING ALGORITHM

In order to build our algorithm, a light propagation analogy is used. In the physical framework, light propagates in space under *Decartes laws* (see Figure 1).

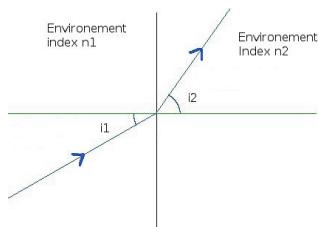


Fig. 1. Light ray deviation under Decartes laws in a region cut into two areas with different refraction indices (n_1, n_2)

These laws are summarized by the following:

For a light ray that goes from a region with index n_1 into an other one with index n_2 with an angle i_1 , we have

$$\begin{cases} n_1 \sin(i_1) = n_2 \sin(i_2) \\ v = \frac{c}{n^2} \end{cases}$$

where, i_2 is the angle of the light ray in region with index n_2 , v its associated velocity and c the velocity of the light in vacuum.

Our algorithm uses such laws in order to build aircraft trajectories using congestion or other aircraft neighborhood as high index areas.

The light modelling algorithm is adjusted from the aircraft point of view. It is assumed that the aircraft knows the surrounding aircraft trajectories (trajectories of other aircraft is a given input of our algorithm).

Assume that our objective function is an application associating a positive real value to a curve of class C^1 of \mathbb{R}^3 . Such a value is computed by integrating a local metric along the curve. We can thus represent length, travel time or the cost associated with a trajectory by a suitable choice of local metric. Determining an optimal trajectory will therefore reduce to search a geodesic which is the shortest path between two points with respect to the local metric.

In this algorithm, we use the well-known fact that a light ray trajectory is a geodesic when considering the environment refractive index as a local metric. To represent congestion areas and conflicts in air traffic management, we consider the refractive index as a measure of congestion or traffic

complexity. We select a barrier index value in the prohibited areas and in the protection volumes surrounding each aircraft. The optimal trajectory will be computed using a technique of ray tracing. The light will be slowed down in congested areas, but despite this, it can pass through. However, it will be completely blocked by aircraft protection volumes, which ensures conflict free-situations. We launch several light rays in various directions from the departure point of the aircraft. The path of the first ray that reaches the arrival point corresponds to an approximation of a geodesic.

We compute the environment index associated to a given congested area using a model based on Lyapunov exponents [3].

To generate a trajectory, we use a wavefront propagation algorithm in 3D with a space discretization (the wave propagation is done with a space step ds) from the departure point. We do not propagate the wavefront randomly in all space directions but into directions with highest probability of success. To ensure this, we guide the wavefront by an initial solution obtained by the navigation function method [2]. Consequently, we can guarantee at least one feasible solution. To avoid a combinatorial explosion, the propagation will be coupled with a branch-and-bound algorithm that interrupts unnecessary shooting rays. In our case, the trajectory obtained by the navigation function method is sampled with half-spheres of radius ds , oriented towards the destination. These half-spheres are, in turn, sampled with an angle step $d\theta$ in the horizontal plane and an angle step $d\varphi$ in the vertical plane to build an initial search tree for the branch-and-bound algorithm as shown in Figure 2.

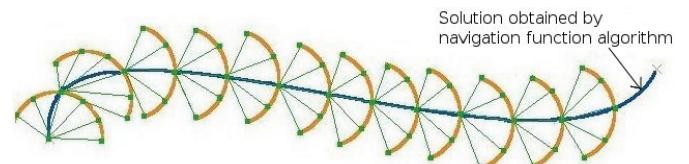


Fig. 2. Part of the initial search tree for the branch-and-bound algorithm

A. Branch-and-bound Algorithm

We implement the propagation within a *branch-and-bound algorithm* [4], a classical framework for solving discrete optimization problems. At the beginning, we consider the set of all possible solutions, represented by the root. Procedures to obtain lower and upper bounds for the optimal value of our criterion are applied to the root. If these two bounds are equals, then the optimal solution is found, and the algorithm stops. Otherwise, the solution set is partitioned into two or more sub-problems, which become children of the root. The method is then applied recursively on these sub-problems, generating a tree.

The idea behind the building of sub-problems which are relaxations of the original problem, branching process, is to solve them in a reasonable time. If an optimal solution is found for a sub-problem, it is feasible but not necessarily optimal for

the original problem. On the other hand, as a feasible solution, it can be used to eliminate partial solutions. Indeed, if the lower bound of a node exceeds the value of an already known solution, then we can say that global optimum solution cannot belong to the subset of solution represented by this node. It is therefore eliminated. The search goes on until all the nodes are explored or eliminated.

B. Branch and Bound applied to the Light Modeling Algorithm

The initial upper bound used in the Branch and Bound algorithm is the travel time computed by the navigation function method.

In order to build a lower bound for a given search tree node, we first compute a bound, "TimeToDest", for the remaining time to reach the destination. This bound is a weighted sum of two terms. The first one, "integTime", is the summation of the refractive index along the direct route to the destination. The second one, "maxSpeedTime", is the time needed to reach destination in direct route with the maximum speed.

$$\text{TimeToDest} := \alpha * \text{integTime} + \beta * \text{maxSpeedTime}. \quad (1)$$

with weighting parameters α, β such that $\alpha + \beta = 1$.

The lower bound is then the summation of TimeToDest and the time needed to reach the node from the origin (TimeToNode). More specifically, the lower bound is given by (see Figure 3):

$$\text{lowerBound} := \text{TimeToNode} + \text{TimeToDest}.$$

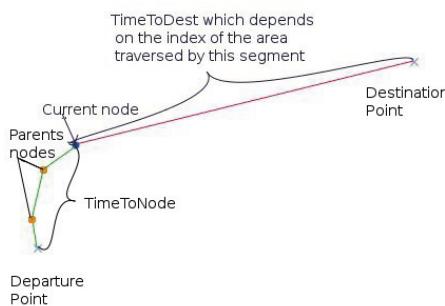


Fig. 3. The lower bound computing.

Branching, in our context, involves launching rays as straight lines in a spatial half sphere of radius ds oriented towards the arrival point.

Browsing the search tree can be done in different ways. We choose a strategy whose priority is to find quickly a feasible solution (depth-first search or DFS). Here a live node (a node for which children have not yet been generated) with deepest level in the search tree is chosen for exploration. The memory requirement in terms of number of subproblems (stored at any given time) is bounded above by the number of levels in the search tree multiplied by the maximum number of children of any node, which is in our context a manageable number.

The drawback of such an approach is that nodes which are far from being optimal, may yield large amount of unnecessary bounding computations.

In order to avoid such a drawback in our case, DFS is combined with a selection strategy. This consists in selecting the node that has the best lower bound among the nodes at the same level in the search tree (a combination of DFS as the overall principle and best first search, BeFS as a secondary selection criterion).

At each time step dt of the algorithm, the environment index is updated because aircraft change themselves the congestion while moving.

The main steps of the algorithm are as follows:

1. Compute the navigation function algorithm solution: trajectory T. Set UpperBound := time travel of T.
2. Descretize T to build a tree, with half spheres having radius ds , an angle step $d\theta$ horizontally and an angle step $d\varphi$ vertically.
3. Set TrajSolution := null. While the destination is not reached do:
 - a. Update the refractive index.
 - b. While there is still unexplored nodes in the tree do:
 - Choose a node N.
 - Relaunch rays from node N:
 - For any light ray, if the light beam goes from a region with index n_1 into a region with index n_2 with an angle i_1 , let it continue with a new angle i_2 such that $n_1 \sin(i_1) = n_2 \sin(i_2)$ and with a velocity of $v = \frac{c}{n_2}$.
 - c. Set TrajSolution := Trajectory obtained at b. interrupted by the step time dt .

In the following section, we will see the numerical results on a simplified instance of the problem (2D), first with a static refractive index and then with a dynamic refractive index.

III. NUMERICAL RESULTS

In all our simulations, we work on a 3.2 GHz machine running under Debian Linux operating system, 1024 KB of RAM. The software has been developed in JAVA.

We use a coordinate system that is scaled with separation standards. Thus, we use an (x, y) grid with a standard horizontal separation (5 Nm) unit. The index map used is a square of $(15 * 15)$ standard horizontal separation.

In step 2. of the algorithm, the radius ds of the semicircles is set to a half standard separation distance. And the sampling angle $d\theta$ is set to $\frac{\pi}{10}$. The coefficients in the formula (1) we chose, are: $\beta := 0.1$ and $\alpha := 0.9$.

A. Results in 2D

The algorithm was first tested with a static refractive index function (it does not depend on time) in 2D space in order to highlight the fact that it does find geodesics in simple cases.

Several refractive index functions were tested. Some examples are presented in Figures 4, 5, 6 and 7. The index

function used is a continuous function, where high values are represented in red and low values in blue. Thus, the congested areas are represented in red and areas that involve little traffic are shown in blue. The resulting solution trajectory is found in less than 5 s of CPU.

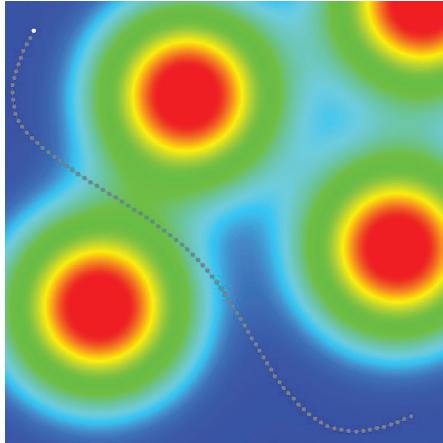


Fig. 4. Resulting trajectory with departure point on the bottom right corner and arrival point on the top left corner. The function index is given by: $e^{-((x-a_1)^2+(y-b_1)^2)/k} + e^{-((x-a_2)^2+(y-b_2)^2)/k} + e^{-((x-a_3)^2+(y-b_3)^2)/k} + e^{-((x-a_4)^2+(y-b_4)^2)/k}$

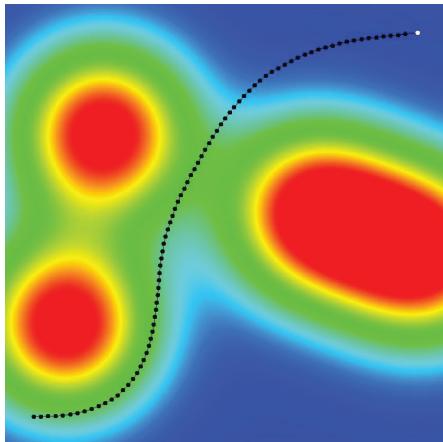


Fig. 5. Resulting trajectory with departure point on the bottom left corner and arrival point on the top right corner. The function index is given by: $e^{-((x-a_1)^2+(y-b_1)^2)/k} + e^{-((x-a_2)^2+(y-b_2)^2)/k} + e^{-((x-a_3)^2+(y-b_3)^2)/k} + e^{-((x-a_4)^2+(y-b_4)^2)/k}$

In Figures 4, 5 and 6, trajectories avoid high index area and pass through "valleys" which is the expected behavior. Thus, the aircraft avoids congested areas.

In Figure 7, the trajectory goes through a relatively congested area instead of bypassing it completely through the blue area above. This behavior can be explained by the fact that direct path, although it slows down the aircraft is more advantageous than a long detour with a higher aircraft velocity.

In all cases, the resulting trajectory is a geodesic approximation.

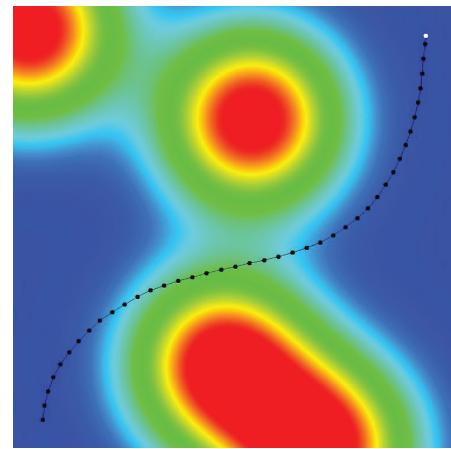


Fig. 6. Idem Figure 5 with new parameters ($a_i, b_i, ii \in \{1..4\}$) for the index function.

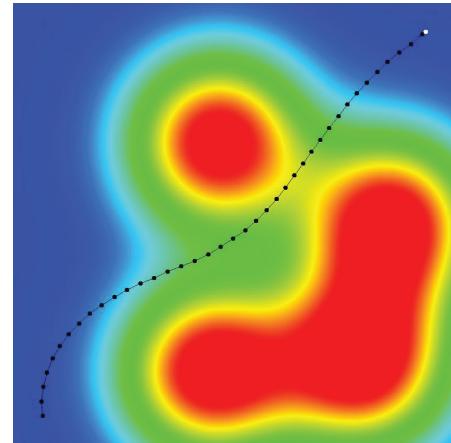


Fig. 7. Idem Figure 5 with new parameters ($a_i, b_i, ii \in \{1..4\}$) for the index function.

B. Results in 2D + time

Here, we deal with a situation of conflict resolution involving several aircraft. The algorithm controls one aircraft trajectory and we assume that the other aircraft maintain fixed straight line trajectories with a 450 knots velocity. The aircraft controlled by the algorithm has an initial velocity of 450 knots, then it varies in a range of [400, 540] knots.

The refractive index takes into account two factors. The first one is related to avoidance of the protection zones of other aircraft. The second one drives away the aircraft from congested areas where it could be in conflict.

The index function n is given by the following formula:

Consider P aircraft $(a_i)_{i \in \{1..P\}}$ moving in the space with the associated position vectors $(\vec{X}_i)_{i \in \{1..P\}}$ and velocity vectors $(\vec{V}_i)_{i \in \{1..P\}}$. For any space point \vec{Y} ,

$$n(\vec{Y}) := n_1(\vec{Y}) + C(\vec{Y}).$$

The first function n_1 is used to avoid conflict. In order to

ensure that the aircraft controlled by the algorithm avoids the other aircraft, we represent them by disks whose radius is the standard distance separation. We set the function n_1 to a very high value N inside these disks and we make it decrease rapidly outside the disk as follows:

For any space point \vec{Y} , let

$$\|\vec{X}_i - \vec{Y}\| := \alpha.$$

$$\begin{cases} \alpha \leq R & \Rightarrow n_1(\vec{Y}) := N \\ \alpha \geq R & \Rightarrow n_1(\vec{Y}) := 1 + \frac{N-1}{1+(\alpha-R)^q}. \end{cases} \quad N \gg 1.$$

with R the standard distance separation and q is a parameter that determines the speed with which the index decreases outside the separation zone.

The second function C called *convergence metric* [5] models the fact that aircraft converge to a point in the space. It is a metric used to measure congestion. The *convergence value* at a space point \vec{Y} is given by the formula:

$$C(\vec{Y}) := \sum_i c(i) * e^{-\beta * (\|\vec{X}_i - \vec{Y}\|)},$$

where

$$c(i) := \sum_{j \neq i, r_{ij} < 0} r_{ij} * e^{-\alpha * (\|\vec{X}_i - \vec{X}_j\|)}$$

where

$$r_{ij} := \frac{(\vec{X}_i - \vec{X}_j)}{\|\vec{X}_i - \vec{X}_j\|} * (\vec{V}_i - \vec{V}_j)$$

and where α and β are weighting parameters.

Some examples of conflict resolutions with 2 to 4 aircraft are shown in Figures 8, 9, 10, 11 and 12.

The track of the aircraft controlled by the algorithm appears in blue. This aircraft has its starting point at the bottom right corner of the figure and its arrival point at the top left corner. Its initial trajectory (a straight line) is shown in red. The other aircraft trajectories appear in black. The starting point of each trajectory is represented by a small dot. Their separation zones are represented by a red disk.

In all these cases, the solution is found in less than 30 s of CPU.

In Figure 8, only two aircraft are in conflict. Their trajectories cross at the center of the figure. The controlled aircraft avoids conflict and follows a path that consists in two segments and an arc.

In Figures 9, 10 and 11, three aircraft are involved. The controlled aircraft is initially in conflict with one or the two other aircraft. And each time, the trajectory generated by the algorithm avoids conflicts while remaining close enough to the original trajectory. Moreover, these trajectories are composed of a set of segments and arcs.

In Figure 12, four planes are considered. As previously, the resulting trajectory avoids conflicts by using 3 segments and an arc.

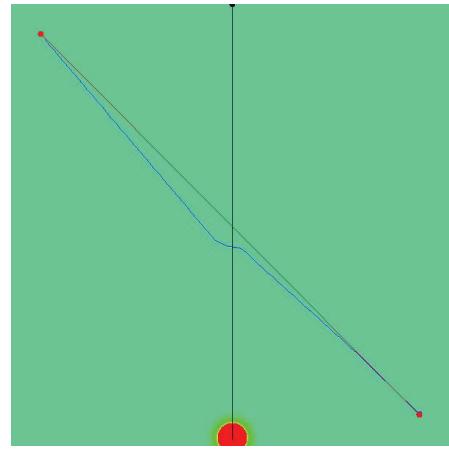


Fig. 8. Conflict resolution with 2 aircraft. The conflicting aircraft trajectory (not controlled by the algorithm) starts on the top in the middle and goes to the bottom.

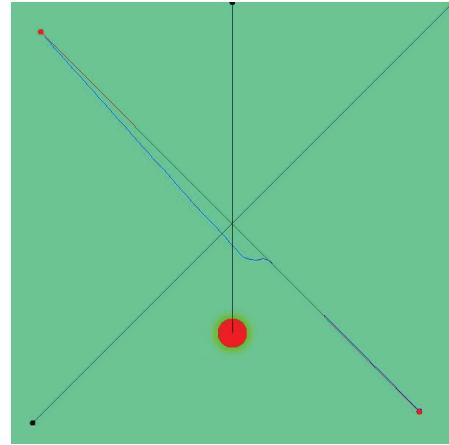


Fig. 9. Conflict resolution with 3 aircraft. The first conflicting aircraft trajectory starts on the top in the middle and goes to the bottom. The second conflicting aircraft trajectory starts on the left bottom corner and goes to the right top corner.

Figure 13 represents a situation with four aircraft in conflict. The resulting trajectory successfully avoids conflicts, but the turns used are too steep. This occurs because the metric used for the index (convergence indicator) does not have a big enough predictability horizon. This creates paths that are not regular enough. Future work will therefore seek for a better congestion metric that avoids the occurrence of such irregularities.

IV. CONCLUSION

The trajectory produced with our light-model algorithm avoids conflicts. It is a geodesic approximation that guarantees a speed lower bound, which is critical for an aircraft. Moreover, this trajectory is a sequence of segments and arcs, a trajectory that the FMS can monitor. Indeed, in most cases, it is a smooth curve which is flyable and which apparently does not require a lot of RTA points (Required Time on Arrival -

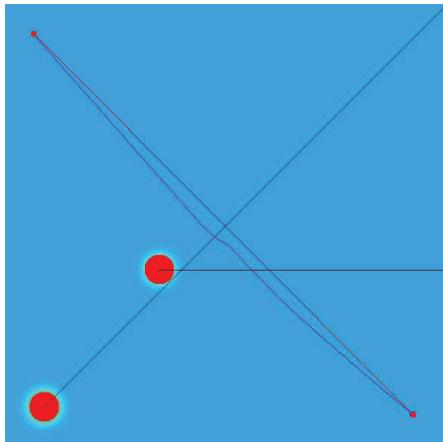


Fig. 10. Conflict resolution with 3 aircraft. The first conflicting aircraft trajectory starts on the right and goes to the left. The second conflicting aircraft trajectory starts on the right top corner and goes to the left bottom corner.

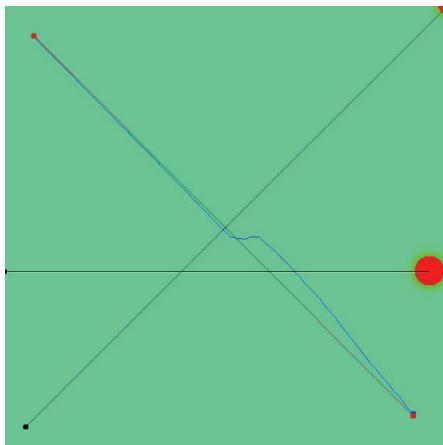


Fig. 11. Conflict resolution with 3 aircraft. The first conflicting aircraft trajectory starts on the left and goes to the right. The second conflicting aircraft trajectory starts on the left bottom corner and goes to the right top corner.

this means that the plane must be in a given position at a given moment). Indeed, RTA points correspond to curvatures changes and in the tested examples, few curvatures changes appear.

In future work, a better convergence indicator will be used. The curvature of the resulting trajectories will be analyzed in order to verify that it satisfies aircraft constraints. This approach allows us to know how many RTA points are required to monitor conflict-free trajectories. Current work involves testing a new method for generating geodesic curves on a triangular mesh following ideas from computer graphics[6]. Our aim is to improve our results both in term of geodesic approximation and in term of computational time.

On the other hand, our algorithm has only been tested in (2D + time) context. Indeed, from the operational point of view, a resolution in the same Flight Level (heading and/or velocity

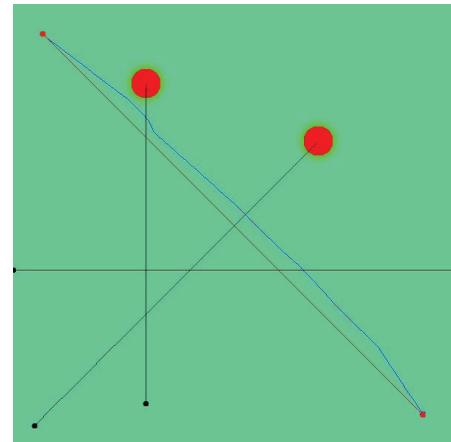


Fig. 12. Conflict resolution with 4 aircraft. The first conflicting aircraft trajectory starts on the left and goes to the left. The second conflicting aircraft trajectory starts on the left bottom corner and goes to the right top corner. The third conflicting aircraft trajectory starts from the bottom and goes to the top.

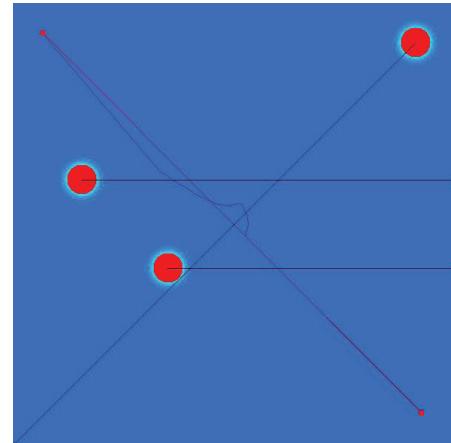


Fig. 13. Conflict resolution with 4 aircraft. The first conflicting aircraft trajectory starts on the left bottom corner and goes to the right top corner. The two other conflicting aircraft trajectory starts from the right and goes to the left.

changes) is always preferable to a resolution with altitude changes. In the future, we will test the algorithm in (3D + time) context. This algorithm version will only be applied to conflicts lefted by the version (2D plus Time) of the algorithm. Indeed, resolution with altitude changes will only be applied if there is no conflict resolution in the same Flight Level.

REFERENCES

- [1] N. Durand. *Optimisation de trajectoires pour la résolution de conflits en route*. PhD thesis, ENSEEIHT, Institut National Polytechnique de Toulouse, France, 1996.
- [2] G. Rousos, G. Chaloulos, K. Kyriakopoulos and J. Lygeros. *Control of multiple non-holonomic air vehicles under wind uncertainty using model predictive control and decentralized navigation function*, IEEE Conference on Decision and Control, December, 2008.
- [3] S. Puechmorel and D. Delahaye. *Dynamical systems complexity with a view towards air traffic management applications*, IEEE Conference on Decision and Control, 2009.
- [4] E. Balas and P. Toth. *Branch and Bound Methods, In The Traveling Salesman Problem*, John Wiley & Sons, 361-401, 1985.
- [5] D. Delahaye, S. Puechmorel. *Air traffic complexity: towards intrinsic metrics*, 3rd USA/Europe Air Traffic Management R& D Seminar Napoli, 2000.
- [6] M. Novotni, R. Klein. *Computing geodisic distances on triangular meshes*, The 10th International Conference in Central Europe on Computer Graphics, 2002.

Resource Allocation in Flow-Constrained Areas with Stochastic Termination Times And Deterministic Movement

Moein Ganji and David J. Lovell

Department of Civil and Environmental Engineering and
Institute for Systems Research
University of Maryland
College Park, USA
moein_g@yahoo.com, lovell@umd.edu

Michael O. Ball

R.H. Smith School of Business and
Institute for Systems Research
University of Maryland
College Park, USA
mball@rhsmsmith.umd.edu

Abstract— In this paper we address a stochastic air traffic flow management problem. Our problem arises when airspace congestion is predicted, usually because of a weather disturbance, so that the number of flights passing through a volume of airspace (flow constrained area – FCA) must be reduced. We formulate an optimization model for the assignment of dispositions to flights whose preferred flight plans pass through an FCA. For each flight, the disposition can be either to depart as scheduled but via a secondary route, or to use the originally intended route but to depart with a controlled (adjusted) departure time and accompanying ground delay. We model the possibility that the capacity of the FCA increases at some future time once the weather activity clears. The model is a two-stage stochastic program that represents the time of this capacity windfall as a random variable, and determines expected costs given a second-stage decision, conditioning on that time. This paper extends our earlier work on this problem by allowing the FCA to move in a 2-D spatial plane with a constant speed rather than being stationary. The FCA can have any given constant speed and any given direction. We conduct experiments considering a range of such speeds and directions and draw conclusions regarding appropriate strategies.

Keywords: ATM; Air Traffic Manegemnt; FCA; Flow Constraint Area; Rerouting; Stochastic Programing; Ground Delay; Airborne Delay.

I. INTRODUCTION

A flow-constrained area (FCA) is a region of the national airspace system (NAS) where a capacity-demand imbalance is expected, due to some unexpected condition such as adverse weather, security concerns, special-use airspace, or others. FCAs might be drawn as polygons in a two-dimensional space, although in practice they are usually represented by a single straight line, functioning as a cordon.

When an FCA has been defined, it is then often the case that an airspace flow program (AFP) is invoked by the Federal Aviation Administration (FAA). An AFP is a traffic management initiative (TMI) issued by the FAA to resolve the anticipated capacity-demand imbalance associated with the FCA. It is the goal of this paper to develop a method by which, given the aggregate data described here, specific orders for individual flights can be developed for a single moving

FCA that a) maximize the utilization of the constrained airspace, b) prevent the capacity of the FCA from being exceeded, and c) achieve a system-wide delay minimization objective. We will emphasize analyzing the effects of a moving FCA due to wind on our model results and will present a methodology to take into account such effects through our model. As reported in weather forecasts, a thunder storm can move up to 50 miles per hour. When the FCA is moving, therefore, a flight that departs a few hours after the beginning of the time horizon may intersect the FCA at a totally different time or location than what would have been calculated for a stationary FCA.

These assumptions are the basis for our motivation to conduct research to investigate the effect of a moving FCA on our model results.

II. RELATED RESEARCH

The research in this paper and our earlier work on this problem builds on stochastic ground holding models. Several stochastic integer programming models have been developed to address the ground holding problem [1], [2], [3], [6], [7], [12]. While our model of the FCA capacity is conceptually similar to airport arrival capacity models, we also explicitly represent the possibility of reroutes, including their dynamic adjustment under stochastic changes in FCA capacity.

There is also a growing literature on airspace flow management problems. Our work also builds on earlier work by Nilim and his coauthors on the use of “hybrid” routes that hedge against airspace capacity changes. In [11], the rerouting of a single aircraft to avoid multiple storms and minimize the expected delay was examined. In this model, the weather uncertainty was treated as a two-state Markov chain, with the weather being stationary in location and either existing or not existing at each phase in time. A dynamic programming approach was used to solve the routing of the aircraft through a gridded airspace, and the aircraft was allowed to hedge by taking a path towards a storm with the possibility that the storm may resolve by the time the aircraft arrived. The focus of the work was on finding the optimal geometrical flight path of the aircraft, and not on allocation of time slots through the weather area. Follow-on work expanded to modeling multiple

aircraft with multiple states of weather and attempted to consider capacity and separation constraints at the storms [9], [10].

Initial steps at a concept of operations that describes the terminology, process, and technologies required to increase the effectiveness of uncertain weather information and the use of a probabilistic decision tree to model the state space of the weather scenarios was provided in [1]. Making use of this framework is a model recently proposed that uses a decision-tree approach with two-stage stochastic linear programming with recourse to apportion flows of aircraft over multiple routing options in the presence of uncertain weather [4]. In the model, an initial decision is made to assign flights to various paths to hedge against imperfect knowledge of weather conditions, and the decision is later revised using deterministic weather information at staging nodes on these network paths that are close enough to the weather that the upcoming weather activity is assumed known with perfect knowledge. Since this is a linear programming model, only continuous proportions of traffic flow can be obtained at an aggregate level, and not decisions on which individual flights should be sent and when they should arrive at the weather. In [8], a stochastic integer programming model is developed based on the use of scenario trees to addressed combined ground delay-rerouting strategies in response to en route weather events. While this model is conceptually more general than ours, by developing a more structured approach we hope to develop a more scalable model.

Recently, a Ration-by-Distance (RBD) method was proposed as an alternative to the Ration-by-Schedule (RBS) method currently used for Ground Delay Programs (GDPs) [5], that maximizes expected throughput into an airport and minimizes total delay if the GDP cancels earlier than anticipated. This approach considers probabilities of scenarios of GDP cancellation times and assigns a greater proportion of delays to shorter-haul flights such that when the GDP clears and all flights are allowed to depart unrestricted, the aircraft are in such positions that the expected total delay can be minimized. While this problem was applied to GDPs, the principles of a probabilistic clearing time where there is a sudden increase in capacity and making initial decisions such that the aircraft are positioned to take the most advantage of the clearing is similar to our problem.

III. THE MODEL

A. Model Inputs

Our base model inputs consist of information about the FCA, which is consistent with the information used in AFP planning:

- Location of the FCA
- Speed and direction of the FCA
- Nominal capacity of the FCA
- Reduced capacity of the FCA
- Start time of the AFP

• Planned end time of the AFP

From a list of scheduled flights and their flight plans, we determine the set of flights whose paths cross the FCA and which therefore would be subject to departure time and/or route controls under an AFP. We also require a set of alternate routes for each flight. The alternate route for each flight should be dependent on the geometry of the FCA and the origin-destination pair it serves. These most likely would be submitted by carriers in response to an AFP; for the purposes of this paper it is assumed they are submitted exogenously, although for testing purposes it was necessary to synthesize alternate routes.

B. Controls

In order not to exceed the (reduced) FCA capacity, each flight will be assigned one of two dispositions in the initial plan reacting to the FCA:

1. *The flight is assigned to its primary route, with a controlled departure time that is no earlier than its scheduled departure time.* Given an estimate of en route time, this is tantamount to an appointment (i.e., a slot) at the FCA boundary. Some flights might be important enough that they are allowed to depart on time, the AFP notwithstanding. Other flights might be assigned some ground delay.
2. *The flight is assigned to its secondary route, and is assumed to depart at its scheduled departure time.*

Several assumptions underlie our model:

- We do not consider airborne holding as a metering mechanism to synchronize a flight on its primary route with its slot time at the FCA.
- We assume that any necessary number of flights can be assigned to their secondary routes without exceeding any capacity constraints in other parts of the airspace.
- We assume that, when the weather clears, the FCA capacity increases immediately, back to the nominal capacity.
- The random variable is the time at which the FCA capacity increases back to its nominal value. We assume that perfect knowledge of the realization of this random variable is not gained until the scenario actually occurs, and so no recourse can be taken until the scenario is realized.

C. Scenarios and future responses

The outputs of this model are:

1. An initial plan that designates whether a flight is assigned to its primary route or secondary route; for those assigned to their primary route an amount of ground delay (possibly zero) is assigned. For those assigned to their secondary route a specific directional angle is assigned.
2. A recourse action for each flight under each possible early clearance time.

We model the time at which the weather clears (i.e. FCA capacity increases) as a discrete random variable, with some exogenous distribution. For any realization of the capacity increase time, the flights in question will be in some particular configuration as specified in the initial plan. Some will have departed, either on their primary or secondary routes, some will already have completed their journeys, and some will still be at their departure airports.

Flights that were originally assigned to their primary route and that have already taken off will be assumed to continue with that plan. For any such flight, the primary route is assumed to be best, so no recourse action is necessary.

We now consider flights originally assigned to their primary route that have not yet taken off. The only possible change in disposition for these flights involves potentially changing their controlled departure time, i.e. reducing their assigned ground delay.

All other flights not yet considered were originally assigned to their secondary routes, with departure times as originally scheduled. These secondary routes avoid the FCA somehow. Under the FCA capacity windfall, some of those flights may now have an opportunity to use the FCA. If a flight has not yet taken off, and it is decided that it can use the FCA, the lowest cost way to do this is to re-assign it back to their primary route, with some controlled departure time no earlier than their scheduled departure time. If, on the other hand, the flight has already taken off, then the only mechanism to allow it the use of the FCA is a hybrid route that includes that portion (and perhaps more) of the secondary route already flown, plus a deviation that traverses the FCA and presumably rejoins the primary route at some point after the FCA (see Fig. 1). A flight that is already en route via its secondary route may or may not prefer such a hybrid path, depending on the difference in cost (time, fuel, etc.) between doing that and continuing on its secondary route. There may be many possible hybrid routes, and perhaps only a limited set of those would be reasonable.

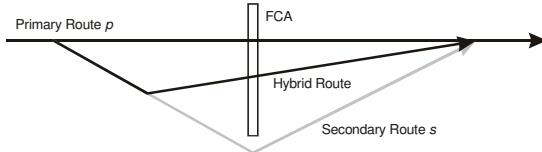


Figure 1. Reverting from secondary route back to primary route through FCA.

For each possible value of the capacity windfall time, we determine the expected locations of all affected flights at that time, and also what would be the best change in disposition, if any, for each of those flights according to a system performance metric. With this information, we can compute the conditional cost associated with these adjusted flights based on the realization of the stochastic event. Ultimately, then, the goal of the optimization problem is to minimize the expected total cost, given these conditional costs and their associated probabilities.

D. Model Development

We start by defining the discrete lattice on which time will be represented. We assume there is an index set $\{1, \dots, T\}$ of

size T that demarcates equally spaced time slots, each of duration Δt . Each of these represents a possible appointment time window at the FCA. The nominal capacity of the FCA should be specified in terms of the maximum number of flights permissible during one of these time windows. The number of time slots T then depends directly on Δt and the total duration of an AFP, perhaps inflated to allow for ending times later than the original estimate. The reference time $t=1$ can be chosen as the earliest scheduled departure time of all of the affected flights.

The flights affected by the FCA can be determined from the filed flight plans for that day, minus known cancellations and re-routes at the time the AFP is invoked. These flights are indexed according to the set $\{1, \dots, F\}$. In this, any specific reference to a time period t and flight f assumes that $t \in \{1, 2, \dots, T\}$ and $f \in \{1, \dots, F\}$.

1) Initial Plan

There are two sets of assignment variables that are related to decisions about the dispositions of flights. One set represents the initial plan, which is the set of decisions provided by the model that will be enacted immediately once the model is run and the AFP is declared. The second set represents conditional decisions (recourse actions) based on the random variable representing the time at which the capacity windfall takes place, which we do not know at the time of the execution of this optimization problem, but that we condition for when determining the best initial plan.

For the initial plan, we define the following set of binary decision variables:

$$x_{f,t}^p = \begin{cases} 1, & \text{if flight } f \text{ uses its primary route and} \\ & \text{has an appointment time } t \text{ at the FCA} \\ 0, & \text{otherwise} \end{cases}$$

$$x_f^s = \begin{cases} 1, & \text{if flight } f \text{ is assigned to its secondary} \\ & \text{route} \\ 0, & \text{otherwise} \end{cases}$$

Every flight f needs to have an assigned disposition under the initial plan, thus:

$$\sum_t x_{f,t}^p + \sum_r x_{f,r}^s = 1 \quad \forall f \quad (1)$$

We require that any flight that is assigned to its primary route cannot be given an appointment slot at the FCA that is earlier than its scheduled departure time plus the expected en route time required to arrive at the FCA. If $E_f \Delta t$ represents the en route time (from its origin to the FCA) for flight f , and $D_f \Delta t$ is the scheduled departure time for flight f , then:

$$\sum_{t=1}^{D_f + E_f} x_{f,t}^p = 0 \quad \forall f \quad (2)$$

No similar constraint is applied to flights assigned to their secondary routes under the initial plan, because they are not metered at any point and hence are expected to depart at their

originally scheduled departure time. There is no provision in the model for a flight to depart early, despite the fact that the secondary route takes more time than the primary route (since, subject to minor variations, airlines do not allow flights to take off before their scheduled departure times).

It might be the case that for a particular flight f , there is a latest slot time l_f at the FCA that the carrier who owns that flight would be willing to accept. Slots later than l_f can be prevented via the following constraint:

$$\sum_{t=l_f+1}^T x_{f,t}^p = 0 \quad (3)$$

For any flight for which l_f is not explicitly provided, l_f is the time beyond which the secondary route will be chosen.

The initial constrained capacity (maximum number of flights) for time window t can now be defined as C_t^0 and the constraint to enforce it is:

$$\sum_f x_{f,t}^p \leq C_t^0 \quad \forall t \quad (4)$$

2) Second Stage

The variables and constraints defined so far represent the first stage of the stochastic program. It is assumed that these decisions will be enacted deterministically immediately after the FCA is declared. Next, we describe the second stage of the stochastic program – those variables that represent the conditional decisions we expect would be made if any of a number of possible capacity windfall times happens to come true in the future. We model the time slot at which this occurs as a discrete random variable with domain Ω and probability mass function

$$f_U(u) = \Pr\{U = u\} \quad \forall u \in \Omega$$

Under a capacity windfall, a flight that was originally assigned to its primary route with a controlled departure time might still be given the same general disposition, although its departure time could be moved earlier if that were beneficial to the system goal. We let

$$y_{f,t}^p | u = \begin{cases} 1, & \text{if at the time } U = u \text{ of the capacity windfall,} \\ & \text{flight } f \text{ is assigned to its primary route with} \\ & \text{appointment slot } t \text{ at the FCA} \\ 0, & \text{otherwise} \end{cases}$$

We will (shortly) introduce other variables for the other possible second stage flight dispositions, and we will require that all flights be assigned a disposition under every possible realization of the stochastic event U . For now, we proceed by obviating values of $y_{f,t}^p | u$ that would either be physically infeasible or politically imprudent. Later, structural constraints plus pressure from the objective function will lead to the best possible selection of second stage dispositions for all flights.

First, it is impossible to assign a flight to a slot that would require it to depart before its scheduled departure time:

$$y_{f,t}^p | u = x_{f,t}^p \quad \forall f, u, \quad \forall t \in \{1, \dots, D_f + E_f\} \quad (5)$$

This constraint works with constraint (2) to achieve the required result.

Given the timing U of the capacity windfall, some flights may already have taken off. If they did so via their primary route (with a controlled departure time), then their second stage disposition should match that of the first stage:

$$y_{f,t}^p | u = x_{f,t}^p \quad \forall f, u, \quad \forall t \in \{1, \dots, u + E_f\} \quad (6)$$

A closer look at constraint (6) reveals that it also satisfies an important requirement for flights that have not yet taken off. For any particular flight f and given the capacity windfall time u , the collection of primary stage variables $\{x_{f,t}^p\}_{t=1}^{t=u+E_f}$ will either contain one at exactly one position or it will consist entirely of zeros. In the former case, this means that the flight has already taken off, and that situation has been dealt with. In the latter case, this is indicative of the fact that these slot times are infeasible. Thus, even for flights that have not yet taken off, constraints (2) and (6) insure that they will not be assigned, in the second stage, to their primary routes with slot times that they cannot achieve.

Looking at constraints (5) and (6), it is clear that they can be combined:

$$y_{f,t}^p | u = x_{f,t}^p \quad \forall f, u, \quad \forall t \in \{1, \dots, \max(u, D_f) + E_f\} \quad (7)$$

On the other hand, for flights that already took off via their secondary routes (and therefore at their scheduled departure times), the only possible second stage dispositions are secondary or hybrid routes, so assignments to primary routes for these flights must be prevented:

$$\sum_t y_{f,t}^p | u \leq 1 - \sum_r x_{f,r}^s \quad \forall u, \forall f \ni D_f < u \quad (8)$$

In addition, we will not allow a flight whose controlled departure time is being moved in the face of a capacity windfall to be worse off than it was before this event materialized:

$$y_{f,t}^p | u \leq \sum_{q \geq t} x_{f,q}^p + \sum_r x_{f,r}^s \quad \forall u, f, t \quad (9)$$

Notice that we want to allow for the possibility that flights originally assigned to their secondary routes can revert, under the appropriate circumstances and if the optimization decides this is best, to their primary route if they have not already taken off, which is why the variable $x_{f,r}^s$ appears in (9).

For flights that were originally assigned to the secondary route, the increased capacity at the FCA might allow some of these flights to pass through the FCA and thus improve their flight path by returning to the primary route at some point after the FCA or continuing directly to the destination. For a flight that has not yet departed, the same structure can apply, but the portions of the total flight path spent on the secondary and reverting routes have length zero. We define the second-stage decision variables for this choice as follows:

$$y_{f,t}^h|u = \begin{cases} 1, & \text{if flight } f \text{ was originally assigned to its secondary route, but under capacity clearing time } u \text{ has been assigned an FCA appointment slot } t \\ 0, & \text{otherwise} \end{cases}$$

This decision can only be reached for flights that were originally assigned to their secondary routes:

$$y_{f,t}^h|u \leq x_f^s \quad \forall u, f, t \quad (10)$$

However, we note that the objective function will enforce this behavior implicitly. Such a flight will be on its secondary route, which may be altered to become a hybrid route that passes through the FCA. We need to impose constraints that insure that these flights are only assigned to FCA time slots they can feasibly reach. If a flight diverts from its secondary route to its hybrid route at time t^d there will be an earliest time it can reach the FCA. Fig.1 illustrates the geometry used to compute the parameter used by our model:

$t_{f,t}^d$ is the time at which flight f must alter its secondary route to become a hybrid route that arrives at the FCA at time t .

The following constraint prevents a flight from diverting to its hybrid route before the weather is actually cleared.

$$y_{f,t}^h|u = 0 \quad \forall f, u, \quad \forall t | t_{f,t}^d \leq u \quad (11)$$

The final option possible is that a flight carries out its originally planned secondary route:

$$y_f^s|u = \begin{cases} 1, & \text{if flight } f \text{ was originally assigned to its secondary route, and if, under AFP stop time } u, \text{ that decision remains unchanged} \\ 0 & \text{otherwise} \end{cases}$$

Practically speaking, it would never make sense to assign a flight to its secondary route under the recourse if it had not also been given the same assignment in the initial plan. It might seem, therefore, that the following constraint is necessary:

$$y_f^s|u \leq x_f^s \quad \forall u, f \quad (12)$$

However, it can be seen that the objective function enforces this behavior implicitly. If it were cost-effective to assign a flight to its secondary route under the recourse, it would also be cost-effective to do so under the initial plan.

Constraints (10) and (12) can be combined into a single constraint:

$$y_{f,t}^h|u + y_f^s|u \leq x_f^s \quad \forall u, f, t \quad (13)$$

It would be possible, given the constraints developed so far, to assign a flight to a hybrid route that essentially reverts to the primary route immediately. In other words, this would be an assignment that is tantamount to taking off on the primary route at the scheduled departure time, which is a more logical way to interpret this outcome. Therefore we introduce the following constraint to enforce this behavior:

$$y_{f,D_f+E_f}^h|u = 0 \quad \forall f, u \quad (14)$$

For each time scenario u , every flight f must be assigned to one of these dispositions. Furthermore, if the disposition involves being scheduled into a slot appointment at the FCA, no more than one slot can be assigned to a given flight. Given that the decision variables are required to be binary, the following constraint addresses both of these concerns

$$\sum_t y_{f,t}^p|u + \sum_t y_{f,t}^h|u + y_f^s|u = 1 \quad \forall u, f \quad (15)$$

For any value $U = u$, there will be a new capacity profile $C^u(t)$ that agrees with $C^0(t)$ up to time $t = u$, but represents an increase in capacity beyond that point. For example, if $C^0(t)$ had been a constant vector, then $C^u(t)$ could be a step function that makes a jump at time $t = u$. On the other hand, if $C^0(t)$ had been a periodic 0-1 function, then $C^u(t)$ might just have an increased duty cycle after time $t = u$. A wide variety of profiles for $C^u(t)$ are possible; the only real requirements are that it agree with $C^0(t)$ prior to time $t = u$, and that after that time, it supports a higher rate of flow than was possible under the initial plan. The capacity constraint under the scenario $U = u$ can now be written as:

$$\sum_f y_{f,t}^p|u + \sum_f y_{f,t}^h|u \leq C_t^u \quad \forall u, t \quad (16)$$

The last constraint prevents flights, for which the FCA will have moved out of their primary path by the time they get there (i.e. $c_f^s = 0$) from being a candidate to get an appointment slot at the FCA

$$\sum_t x_{f,t}^p \leq M c_f^s \quad \forall f \quad (17)$$

where M is a constant number that makes $M c_f^s > 1$ for all flights f . By this constraint such flights will be forced to be assigned to their secondary route, which will cause no delay.

3) Objective Function

Since our model involves the specification of decisions that are conditioned random events, the objective function will be an expected value. To emphasize the paradigm of creating a plan (our initial plan) together with contingency plans (our recourse actions), we represent the objective function as the sum of the deterministic cost of the initial plan minus the expected savings from recourse actions.

Therefore the objective function can thus be represented as:

$$\text{Min} \left[C(X) - \sum_u P_u S(Y_u) \right] \quad (18)$$

or more precisely:

$$\text{Min} \quad Z = z^1 + z^2 - \sum_u P_u (z_u^3 + z_u^4) \quad (19)$$

where

$$z^1 = \sum_f \sum_t c_{f,t}^p x_{f,t}^p \quad (20)$$

$$z^2 = \sum_f c_f^s x_f^s \quad (21)$$

$$z_u^3 = z^1 - \sum_f \sum_t c_{f,t}^p y_{f,t}^p |u + \sum_f \sum_t c_f^s s_{f,t}^p |u \quad (22)$$

$$z_u^4 = \sum_f \sum_t sv_{f,t}^h y_{f,t}^h |u \quad (23)$$

where

$c_{f,t}^p$ is the cost of assigning flight f to its primary route so that it arrives at the FCA at time t .

c_f^s is the cost of assigning flight f to its secondary route.

$sv_{f,t}^h$ is the savings incurred if flight f starts out on its secondary route but reverts to a hybrid route that arrives at the FCA at time t .

$s_{f,t}^p$ is a dummy binary variable that works as an indicator. It takes the value of one when a flight initially assigned to its secondary route is assigned back to its primary route under the revised plan.

$$s_{f,t}^p = \text{Min}(x_f^s, y_{f,t}^p) \quad \forall f, t \quad (24)$$

IV. THE PARAMETERS

In the following section we will show how the movement of the FCA will affect our previous flight path geometries and we will provide the related calculations for each case. One can obtain the same functions for the case of a stationary FCA simply by setting v_a and v_c to zero.

The first set of equations show how the primary route cost functions are recalculated. We assume that the intersection of the FCA and a flight primary path (i.e. the point that the flight will enter the FCA) will move with a constant speed either toward or away from the flight. In these equations, v_f is the speed of the aircraft along its path, v_a is the projection of the FCA speed vector on the flight's primary path, v_c is the orthogonal component of that velocity, t is the time of arriving at the FCA, and t_d and t_s are the actual and scheduled departure times.

$$\begin{aligned} v_f(t-t_d) + v_a(t-t_d) &= a - v_a(t_d - t_s) \quad \forall t \\ t_d &= t + \frac{v_a}{v_f}(t-t_s) - \frac{a}{v_f} \end{aligned} \quad (25)$$

$$c_{f,t}^p = t_d - t_s = \left(1 + \frac{v_a}{v_f}\right)(t-t_s) - \frac{a}{v_f} \quad \text{and } c_{f,t}^p \geq 0$$

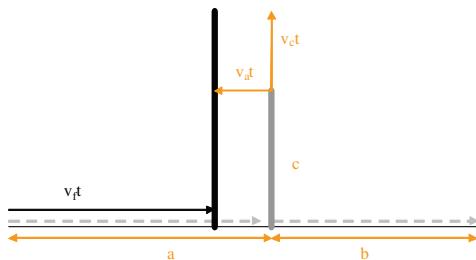


Figure 2. Primary route

Next we show how the secondary cost functions are recalculated. Let α and β be the required directional angles of the reroute to avoid a moving and stationary FCA, respectively. The gray dashed lines represent the flight path in the case of a stationary FCA and the black lines represent those of a moving FCA.

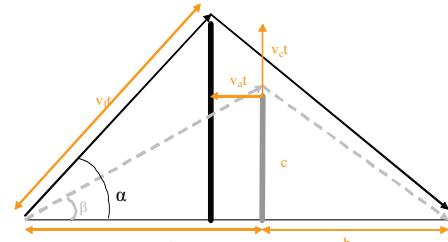


Figure 3. Secondary route

$$\begin{aligned} v_f(t-t_f^d) \cos \alpha &= a - v_a(t-t_f^d) \\ v_f(t-t_f^d) \sin \alpha &= c + v_c(t-t_f^d) \end{aligned} \Rightarrow$$

$$t-t_f^d = \frac{a}{v_f \cos \alpha + v_a} \Rightarrow \frac{av_f \sin \alpha}{v_f \cos \alpha + v_a} = c + \frac{av_c}{v_f \cos \alpha + v_a}$$

$$\Rightarrow a \sin \alpha - c \cos \alpha = \frac{cv_a + av_c}{v_f}$$

$$d = \sqrt{c^2 + a^2}, \cos \beta = \frac{a}{d}, \sin \beta = \frac{c}{d}$$

$$\Rightarrow \frac{a}{d} \sin \alpha - \frac{c}{d} \cos \alpha = \frac{cv_a + av_c}{dv_f} = \cos \beta \sin \alpha - \sin \beta \cos \alpha \Rightarrow$$

$$\sin(\alpha - \beta) = \frac{cv_a + av_c}{dv_f} \quad (26)$$

With the new directional angle α , we can calculate the cost of airborne delay of the reroute;

$$\begin{aligned} d' &= \frac{a}{v_f \cos(\alpha) + v_a} \\ c^s &= d' + \sqrt{d'^2 + (a+b)^2 - 2\cos(\alpha)d'(a+b)} - a - b \end{aligned} \quad (27)$$

Finally, once we have found the adjusting angle for the secondary route compromising the FCA movement, we need to calculate the interrelated changes to our hybrid route cost saving function as well. To do so we assume that the weather clears after i minutes of the flight's departure. With the speed of v_f , our flight would traverse a distance $\overline{AD} = i$ times v_f along its secondary route. As shown in Fig. 4, by knowing \overline{AD} we can compute its counterpart angle μ .

$$\cos \mu = \frac{a+b - iv_f \cos \alpha}{\sqrt{(iv_f)^2 + (a+b)^2 - 2iv_f(a+b)\cos \alpha}} \quad (28)$$

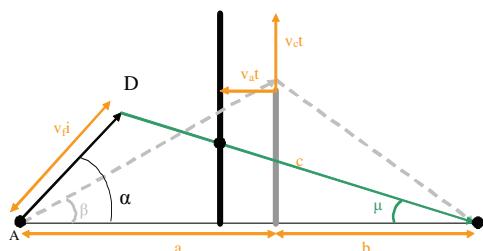


Figure 4. Hybrid route

Now that we have μ , we can build in our governing equation to calculate the time t , at which the flight arrives at FCA if it reverts from its secondary route after i minutes of its departure.

$$v_f i \cos \alpha + v_f (t - i) \cos \mu = a - v_a t \Rightarrow \\ t = \frac{a + iv_f (\cos \mu - \cos \alpha)}{v_f \cos \mu + v_a} \quad (29)$$

This is only true if the flight has not yet reached the end of the FCA (i.e. point C). Therefore the following constraints should apply to maintain the feasibility of the above equations.

$$v_f i \sin \alpha - v_f (t - i) \sin \mu \leq c + v_c t \quad (30)$$

and finally the saving incurred on the hybrid route:

$$sv_{f,t}^h = c_f^s - v_f i_t - \sqrt{(v_f i_t)^2 + (a+b)^2 - 2v_f i_t (a+b) \cos \alpha} \quad (31)$$

V. COMPUTATIONAL EXPERIMENTS

A. Decision Impacts

To evaluate the impact of the FCA movement on our model we ran a set of experiments, where we varied the direction of the FCA movement. In this way we were able to mimic more realistic environments where the flights' paths (primary and reroute) are affected by the movement of the FCA as well as its presence. The cases vary relative to the speed, direction and severity of the weather activity and recourse actions are allowed and planned respectively. A recourse action is taken if the weather clears earlier than expected. In the ground delay case, this means a flight is released at a time earlier than its planned departure time. In the reroute case, this means a flight adjusts its original planned reroute to a more direct route. The key novel contribution of our model is its ability to take into account recourse actions when generating its initial plan.

We now describe the problem data. Flights, their routes, and alternate routes were generated artificially based on the airspace geometry given in Fig. 1. There were $F=200$ flights with random departure times ($t_s=0,\dots,60$). There were $T=200$ time slots; each slot had a width of $\Delta t=2$ minutes. There were three possible early clearance times: $U \in \{30, 50, 70\}$ each occurring with probability 0.3 and 0.1 is the probability that the FCA does not clear until the end time of the AFP. The following alternate cases were considered. The ratio of

airborne delay cost to ground delay cost was assumed to be 2.0.

Case 1: This case considers a stationary FCA and runs the model to find the best initial plan which will serve as a base for the purpose of evaluation of the other cases.

Case 2-9: in these cases the FCA has eight different directions with the same velocity approximately equal to 5% of the average flight speed. The reduced throughput of the FCA is one flight every 4 minutes and increased throughput is 2 flights per minute.

Case 10-17: these cases are similar to cases 2-9 except that the FCA velocity is approximately equal to 10% of the average flight speed.

Case 18: this case is similar to case 1 except that the reduced throughput of the FCA is one flight every 8 minutes and increased throughput is one flight per minute.

Case 19-26: these cases are similar to cases 10-17 except that the reduced throughput of the FCA is one flight every 8 minutes and increased throughput is one flight per minute.

The table below provides the results of an experiment under which all 26 cases were executed. First of all it should be clarified that for simplicity all the 200 flights are assumed to fly in the same direction but with different origin-destination distances, different scheduled departure times and different directional angles for their reroutes. Valuable insights should be obtainable even with this simplification, and more realistic scenarios can always be studied with the exact same formulation.

The first thing to notice is that the movement of the FCA can significantly change the total cost as well as the assignment of the dispositions to all flights affected by the presence of the FCA. The second interesting result is the consistent pattern with which the objective function value increases. In the result table we have sorted the similar cases (similar in terms of the FCA velocity and throughput) in an increasing order of the objective function value. In all three sections of the results table, perhaps not surprisingly, the maximum cost saving occurs when the FCA moves laterally (downward in Fig. 2 and Fig. 3), in which case it either gets out of the way of the primary paths of the affected flights quickly or lowers the maximum length of the reroutes. The total cost is reduced by 38% and 64% with the low and high speed FCA, respectively.

One can observe that the effect of the longitudinal movement of the FCA, where it moves either toward or away from the oncoming traffic, is less significant than the lateral motion. The total cost is increased by 3% (8% for the high speed FCA) when the FCA is moving away from the traffic. When it is moving toward the traffic the total cost is increased by 19% (23% for the high speed FCA).

The second and the third columns are the total costs for ground delays and airborne delays of the first stage. The fourth and the fifth columns are the numbers of flights assigned to primary and secondary paths. The sixth column is the objective function value, which is the minimum expected total cost. The seventh and the eighth columns are the horizontal

TABLE 1. Experimental results on the effects of a moving FCA

Case	c(xp=1)	c(xs=1)	n(xp=1)	n(xs=1)	Obj	Va	Vc	Wind
1	80	459	67	133	379	0.00	0.00	●
2	66	225	60	140	236	0.00	-0.05	↓
3	97	285	65	135	310	-0.035	-0.035	↗
4	103	300	60	140	336	0.035	-0.035	↖
5	124	494	71	129	391	-0.05	0.00	→
6	108	437	61	139	452	0.05	0.00	←
7	130	689	70	130	543	-0.035	0.035	↗
8	117	670	67	133	572	0.035	0.035	↖
9	105	783	70	130	585	0.00	0.05	↑
10	65	93	58	142	135	0.00	-0.10	↓
11	98	161	63	137	185	-0.07	-0.07	↗
12	84	173	56	144	234	0.07	-0.07	↖
13	123	528	74	126	409	-0.10	0.00	→
14	102	451	61	139	466	0.10	0.00	←
15	150	1020	75	125	655	-0.07	0.07	↗
16	130	899	66	134	752	0.07	0.07	↖
17	135	1192	71	129	799	0.00	0.10	↑
18	65	717	36	164	605	0.00	0.00	●
19	55	190	32	168	247	0.00	-0.10	↓
20	85	331	36	164	327	-0.07	-0.07	↗
21	94	305	32	168	389	0.07	-0.07	↖
22	115	892	39	161	635	-0.10	0.00	→
23	90	712	34	166	711	0.10	0.00	←
24	133	1554	41	159	972	-0.07	0.07	↗
25	143	1333	34	166	1082	0.07	0.07	↖
26	107	1812	38	162	1178	0.00	0.10	↑

and the vertical component of the FCA velocity vector. The last column visualizes the FCA direction. The units of all costs are “numbers of time slots,” which can be converted readily to minutes, and presumably to dollars if the analyst has data on economic time values.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we have defined the basics of a stochastic optimization model for simultaneously making ground delay and reroute decisions in response to moving en route airspace congestion. We have also given the results of computational experiments that test the impact of the speed and direction of the movement of the flow-constrained area on decisions as well as the outcome. We believe that the model can serve as a basis for solving practical TFM problems using commercial IP solvers. Further, the results show that the models have the potential to substantially improve TFM decision making.

Our model can be re-run if, and as often as, real-time information suggest that the data supporting a previous execution of the model have changed significantly, for example, if carriers cancel some additional flights, or if the probabilistic weather forecast changes. The model can be

forced to preserve earlier decisions by additional constraints fixing those decisions for flights currently in the air.

We anticipate the need to provide more refinements and extensions to this model to better address practical problem solving.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support of the National Aeronautics and Space Administration Airspace Systems Program under ARMD NRA: NNH06ZNH001.

REFERENCES

- [1] Ball, M. O., R. Hoffman, A. Odoni, and R. Rifkin. “A Stochastic Integer Program with Dual Network Structure and Its Application to the Ground-Holding Problem.” *Operations Research*, Vol. 51, 2003, pp. 167–171.
- [2] Ball, M. O., and G. Lulli. “Ground Delay Programs: Optimizing over the Included Flight Set Based on Distance.” *Air Traffic Control Quarterly*, Vol. 12, 2004, pp. 1–25.
- [3] Davidson, G., and Krozel, J. “Strategic Traffic Flow Management Concept of Operations.” *Aviation Technology, Integration and Operations (ATIO)*, Chicago, IL, 2004.
- [4] Hoffman, R., J. Krozel, G. Davidson, and D. Kierstead. “Probabilistic Scenario-Based Event Planning for Traffic Flow Management.” *AIAA Guidance, Navigation, and Control Conference*, Hilton Head, SC, 2007.
- [5] Ball, M.O., R. Hoffman and A. Mukherjee, “Ground Delay Program Planning under Uncertainty Based on the Ration-by-Distance Principle”, *Transportation Science*, in press, 2009.
- [6] Kotnyek, B., and O. Richetta, “Equitable Models for the Stochastic Ground-Holding Problem Under Collaborative Decision Making.” *Transportation Science*, Vol. 40, 2006, pp. 133–146.
- [7] Mukherjee, A., and M. Hansen. “A Dynamic Stochastic Model for the Single Airport Ground Holding Problem.” *Transportation Science*, Vol. 41, 2007, pp. 444–456.
- [8] Mukherjee, A., and M. Hansen. “A Dynamic Rerouting Model for the Air Traffic Flow Management.” *Transportation Research Part B: Methodological*, Vol. 43, No. 1, 2009, pp. 159–171.
- [9] Nilim, A., and L. El Ghaoui. “Algorithms for Air Traffic Flow Management Under Stochastic Environments.” *IEEE American Control Conference*, Evanston, IL, 2004.
- [10] Nilim, A., L. El Ghaoui, and V. Duong. “Multi-Aircraft Routing and Traffic Flow Management Under Uncertainty.” *USA/Europe Air Traffic Management R&D Seminar*, Budapest, Hungary, 2003.
- [11] Nilim, A., L. El-Ghaoui, V. Duong, and M. Hansen. “Trajectory-Based Air Traffic Management (TB-ATM) Under Weather Uncertainty.” *USA/Europe Air Traffic Management R&D Seminar*, Santa Fe, NM, 2001.
- [12] Richetta, O., and A. R. Odoni. “Solving Optimally the Static Ground Holding Problem in Air Traffic Control.” *Transportation Science*, Vol. 27, 1993, pp. 228–238.

The Air Traffic Flow Management Problem with Time Windows

Luca Corolli and Lorenzo Castelli

Dipartimento di Eletrotecnica,

Elettronica e Informatica

Università degli Studi di Trieste

Via A. Valerio 10, 34127 Trieste, Italy

Email: luca.corolli@deei.units.it, castelli@units.it

Guglielmo Lulli

Dipartimento d'Informatica,

Sistemistica e Comunicazione

Università degli Studi di Milano - Bicocca

Viale Sarca 336, 20126 Milano, Italy

Email: lulli@disco.unimib.it

Abstract—This paper defines a set of temporal intervals, called time windows, which are defined prior to flight departure and constitute milestones to be met during the flight execution. The size of the time windows is variable as it reflects all known constraints, such as punctuality at destination, runway capacities or congested en-route areas that the flight will cross. Once a time window is defined, all the air traffic actors are committed to guarantee that flight operations, e.g. enter an airspace sector, depart from or arrive at an airport, are executed within the time window. We propose a two-step approach based on a mixed integer programming formulation. The first step determines a set of time windows such that the overall cost of delay is minimized. Then in the second step we choose the set of optimal time windows which also maximizes the overall time window size. In such a way, we provide to all air traffic stakeholders the largest degree of flexibility to perform their operations under the constraint that the minimum achievable delay is kept constant. We also gain information on the critical flights of the system: if the optimal width of a time window is equal to its minimum available value, any disruption that may cause the flight not to meet it may produce undesired downstream effects. Our preliminary computational experience based on small-scale random instances confirms that the flexibility granted to flights increases with the capacity while the system delay simultaneously decreases. We also show that when there is no congestion a non negligible share of small size time windows may exist, thus indicating the existence of bottlenecks and critical flights.

Keywords - Air Traffic Flow and Capacity Management, Time Windows, Delay, ATFM, ATFCM

I. INTRODUCTION

The Air Transportation System both in Europe and in the United States is highly capacity constrained due to the limited availability of resources both on the ground and in en-route airspace. Capacity at airports is limited by the runway systems and the terminal airspace around them [1]. The capacity of en-route airspace sectors is limited by the maximum workload acceptable for air traffic controllers [2]. These capacity constraints are becoming a limiting factor in many regions of the world. In fact, as air traffic grows and/or capacity is reduced - mainly due to adverse weather conditions -, demand can exceed capacity at key points of the air transportation network and at critical times. These local overloads create delays which propagate to other parts of the air network, amplifying congestion as increasing number of local capacity constraints

come into play. Air traffic flow management (ATFM) attempts to prevent local demand-capacity imbalances by adjusting the flows of aircraft on a national or regional basis [3]. In Europe, delays caused by ATFM measures in 2007 amounted to 21.5M minutes, producing an estimated cost of M€1300 to airlines [4].

The ATFM problem was first formalized in 1987 by Odoni [5]. Since then, a plethora of mathematical models have been developed, but most of these models focus on airport congestion. Some of them consider extensions to a network of airports. This class of models optimizes the ground delay assignment to various flights, so that delay on a given flight segment can propagate to downstream segments flown by the same aircraft, e.g., see [6] and [7]. In contrast to the case in which solely airport congestion is considered, the research literature dealing simultaneously with airport and en-route congestion is quite sparse. One of the first attempts to include en-route capacity restrictions in the ATFM problem was by Helme [8], who proposed a multi-commodity minimum-cost flow on a time-space network to assign airborne and ground delay to aggregate flows of flights. Lindsay et al. [9] formulated a disaggregate deterministic 0-1 integer programming model for assigning ground and airborne holding to individual flights in the presence of both airport and airspace capacity constraints. Bertsimas and Stock [10] provided a strong formulation of the ATFM problem. In 2007, Lulli and Odoni [11] illustrated the complex nature of the European ATFM system where congestion in the en-route airspace is an issue. They show that counter-intuitive solutions exist when assigning delays to the different phases of the flight. They also discuss the conflicts that may arise between the objectives of efficiency or equity. Recently, Bertsimas, Lulli and Odoni [3] presented an Integer Programming model for the ATFM problem. They provide a complete representation of all the phases of a flight, and modeled a wide set of control actions, including rerouting. Extensions of this work may be found in [12] and [13]. The interested reader may refer to [14] for a detailed relevant survey on ATFM models.

Most if not all of the models developed for the ATFM minimize the delay to be assigned to flights in order to resolve local demand-capacity imbalances. However, none of these

models - to the best of our knowledge - explicitly considers the criticality of the flight. The execution of a flight - from the Air Traffic Control (ATC) point of view - requires a complex mix of capacitated resources which is negotiated between air traffic controllers and flight crews (or dispatchers) according to the air traffic conditions and is concretized in an approved flight plan. The operator of a flight is expected to adhere as precisely as possible to the flight plan, although some adjustments are possible. However, there are flights which have to be operated in strict accordance to the approved flight plan, since any delay assigned to them may have large downstream effects such as disruptions in the airline schedules or degradations of the ATC system performances. For these flights, there is no slack time in handling their operations and a limited number of recovery options is generally available.

Herein, we extend the Bertsimas, Lulli and Odoni model to detect “critical” flights. For each aircraft and for each phase of the flight the model identifies a temporal interval, also called time window (TW), inside which each air traffic actor engages in delivering its services to flight execution, from gate to gate. This allows the definition for each aircraft of a number of TWs located at the transfer of responsibility areas along flight trajectory (e.g. between two sectors) with specific temporal duration, determined according to resource availability (e.g. capacity) and downstream constraints (e.g. punctuality at the destination). In other words, a time window is a period of time associated with a specific phase of the flight, e.g., taking off, landing and entering sectors, which has to be executed within it. This model guarantees that such actions can be executed at any instant within its time window, respecting all the constraints. For any flight the width of the window is not a priori determined as it may depend on other flights’ requests: for instance, several flights may wish to enter the same sector at approximately the same time. Thus the TW size delineates the degree of flexibility to carry out a specific operation. The larger is the time window, the larger is the amount of slack time. The smaller is the time window, the more “critical” is the flight. For critical flights, it is important that all the activities executed in support of flight operations, e.g. maintenance, ground and flight crew activities and ATC clearances, are coordinated and executed on time.

We refer to the model herein presented as the Air Traffic Flow Management problem with Time Windows (ATFMTW). Our formulation lexicographically sets two objectives. The first step identifies the optimal sizes of the time windows to minimize the total cost of delay. Costs are defined by two super-linear functions, one for the departure and the other for the arrival delay costs, which ensure fairness of delay distribution between the various flights [11]. Given the minimum delay cost that can be assigned to the whole system, the second goal is to maximize the total width of the existing time windows. As some portions of the airspace may be less congested than others, a larger flexibility or room of maneuver can be given to airspace users, air navigation service providers and airports operating in such sparse areas without degrading the overall performance (i.e., the total cost of delay) of the

entire system. On the other side, small-sized time windows impose strict limitations on the different elements of the system as any actor is urged to assist flights to respect them as much as possible (critical flights).

This paper unfolds as follows: Section II presents the mathematical formulation of the model. Section III describes the computational experience to date and analyses the results. Finally, Section IV summarizes conclusions and indicates the next research steps.

II. MODEL’S MATHEMATICAL DESCRIPTION

Here we present the mathematical model for the ATFM with Time Windows (ATFMTW). As mentioned above this model can be envisioned as an extension of the model presented in [3]. As such, we define similar decision variables as described below.

For the sake of clarity, we first define the following notation:

- \mathcal{K} ≡ set of airports
- \mathcal{S} ≡ set of sectors
- \mathcal{F} ≡ set of flights
- $f \in \mathcal{F}$ ≡ generic flight
- $\mathcal{S}^f \subseteq (\mathcal{S} \cup \mathcal{K})$ ≡ set of sectors that can be flown by flight f , including the origin and destination airports of f
- \mathcal{T} ≡ set of time periods
- \mathcal{C} ≡ set of pairs of flights that are continued
- \mathcal{P}_i^f ≡ set of sector i ’s preceding sectors
($i \in \mathcal{S}^f$)
- \mathcal{L}_i^f ≡ set of sector i ’s subsequent sectors
($i \in \mathcal{S}^f$)
- $\mathcal{H} = [HI, HF]$ ≡ set of capacity periods
 - HI ≡ initial instant of \mathcal{H}
 - HF ≡ final instant of \mathcal{H}
- NTH ≡ number of time periods in a single capacity period
- $D_k(h)$ ≡ departure capacity of airport k at capacity period h
- $A_k(h)$ ≡ arrival capacity of airport k at capacity period h
- $S_j(h)$ ≡ capacity of sector j at capacity period h
- s_f ≡ turnaround time of an airplane after flight f
- $orig_f$ ≡ airport of departure of flight f
- $dest_f$ ≡ airport of arrival of flight f
- $l_{fjj'}$ ≡ minimum number of time periods that

flight f must spend in sector j before
entering in sector j'

$end_f \equiv$ maximum acceptable duration of flight f

$T_j^f = [\underline{T}_j^f, \bar{T}_j^f] \equiv$ set of feasible time periods for flight f to
to depart from $j = orig_f$ or arrive at
 $j = dest_f$ or enter sector j

$\underline{T}_j^f \equiv$ first time period in the set T_j^f

$\bar{T}_j^f \equiv$ last time period in the set T_j^f

$MINTW \equiv$ the minimum time window size

$MAXTW \equiv$ the maximum time window size

A. Decision variables

As the width of the time windows is not a priori determined, we need to define for each flight f that can cross sector $j \in S^f$ two sets of decision variables: the beginning and the end of the time window. Therefore, we introduce the following binary decision variables:

$$wi_{j,t}^f = \begin{cases} 1, & \text{if time window for flight } f \text{ in sector } j \\ & \text{has been opened by time } t \\ 0, & \text{otherwise} \end{cases}$$

$$wf_{j,t}^f = \begin{cases} 1, & \text{if time window for flight } f \text{ in sector } j \\ & \text{has been closed by time } t \\ 0, & \text{otherwise} \end{cases}$$

For each flight f and sector/airport j , the decision variables do not need to be defined for each $t \in T$ but only on the set of feasible time periods T_j^f . Note that one variable for each flight-airport can be eliminated from the formulation. As flight's cancellation is not considered in the model, we can set variables $wf_{orig_f, \bar{T}_{orig_f}}^f$ and $wf_{dest_f, \bar{T}_{dest_f}}^f$ to 1 for each flight f , since flight f has to leave from airport $orig_f$ and arrive to airport $dest_f$. Finally, we introduce additional decision variables to formulate the constraints associated to the utilization of the available capacity. The sector capacity is defined as the maximum number of flights that can be in a sector during a "capacity period" h composed of a set of contiguous time periods. Then, to determine the capacity occupancy for a flight f of sector j at some capacity period h , we define the following binary decision variables:

$$co_{j,h}^f = \begin{cases} 1, & \text{if flight } f \text{ enters sector } j \text{ during} \\ & \text{capacity period } h \\ 0, & \text{otherwise} \end{cases}$$

We show next that the integrality condition can be relaxed.

B. Objective functions

The first objective is the minimization of the total delay cost. For a flight f this total delay cost is defined as the sum of the departure and arrival delay costs. Thus we introduce two cost coefficients $ddc^f(t)$ and $adc^f(t)$ which represent the

delay cost for flight f when the departure and arrival time window is closed at time t , respectively:

$$ddc^f(t) = \begin{cases} 0 & \text{if } t \leq z_{orig_f} \\ (t - z_{orig_f})^{1+\epsilon_d} & \text{otherwise} \end{cases}$$

$$adc^f(t) = \begin{cases} 0 & \text{if } t \leq z_{dest_f} \\ (t - z_{dest_f})^{1+\epsilon_a} & \text{otherwise} \end{cases}$$

where z_{orig_f} and z_{dest_f} are respectively the last time periods in which flight f can depart from its origin and arrive at its destination airport without causing a delay. If DEL is the maximum allowed delay, it easily follows that $z_{orig_f} = \bar{T}_{orig_f} - DEL$ and $z_{dest_f} = \bar{T}_{dest_f} - DEL$. The values $\epsilon_d > 0$ and $\epsilon_a > 0$ are two positive parameters which make the cost coefficients super-linear. As proposed in [11], this choice grants a fair assignment of delay among different flights.

Therefore, the objective function minimizing the total cost of delay is formulated as follows:

$$\begin{aligned} Z_1 = Min \sum_{f \in \mathcal{F}} \sum_{t \in T_{orig_f}^f} & ((wf_{orig_f,t}^f - wf_{orig_f,t-1}^f) \cdot ddc^f(t)) + \\ & + \sum_{f \in \mathcal{F}} \sum_{t \in T_{dest_f}^f} ((wf_{dest_f,t}^f - wf_{dest_f,t-1}^f) \cdot adc^f(t)) \end{aligned}$$

Once the minimum cost of delay is attained, the second objective is to make the time windows as large as possible. In this way we may provide greater flexibility to the different stakeholders with no harm to the overall system performance. Time windows that cannot be enlarged allow a) to identify critical flights, i.e., flights whose operations need to be performed within the specified time windows, otherwise a delay would occur, and b) to spot the airspace bottlenecks, i.e., the portions of airspace which are most congested.

Mathematically, the second objective function searches its optimal solution on the polyhedron made of the optimal solutions of the first step. Then it maximizes the number of time periods composing each time window:

$$Z_2 = Max \sum_{f \in \mathcal{F}, j \in S^f} \sum_{t \in T_j^f} (wi_{j,t}^f - wf_{j,t-1}^f),$$

being subject to the additional constraint:

$$\begin{aligned} Z_1 = \sum_{f \in \mathcal{F}} \sum_{t \in T_{orig_f}^f} & ((wf_{orig_f,t}^f - wf_{orig_f,t-1}^f) \cdot ddc^f(t)) + \\ & + \sum_{f \in \mathcal{F}} \sum_{t \in T_{dest_f}^f} ((wf_{dest_f,t}^f - wf_{dest_f,t-1}^f) \cdot adc^f(t)) \end{aligned}$$

C. Constraints

The model's constraints set is as follows:

$$\begin{aligned} co_{j,h}^f &\geq wi_{j,\min((h-HI+1)\cdot NTH+HI-1,\bar{T}_j^f)}^f - \\ &- wf_{j,(h-HI)\cdot NTH+HI-1}^f \end{aligned} \quad \forall f \in \mathcal{F}, \forall j \in \mathcal{S}^f, \forall h \in \mathcal{H} \quad (1)$$

$$\sum_{f \in \mathcal{F}: orig_f=k} co_{k,h}^f \leq D_k(h) \quad \forall k \in \mathcal{K}, \forall h \in \mathcal{H} \quad (2)$$

$$\sum_{f \in \mathcal{F}: dest_f=k} co_{k,h}^f \leq A_k(h) \quad \forall k \in \mathcal{K}, \forall h \in \mathcal{H} \quad (3)$$

$$\sum_{f \in \mathcal{F}: j \in \mathcal{S}^f} co_{j,h}^f \leq S_j(h) \quad \forall j \in \mathcal{S}, \forall h \in \mathcal{H} \quad (4)$$

$$wi_{j,t+1}^f - wi_{j,t}^f \geq 0 \quad \forall f \in \mathcal{F}, \forall j \in \mathcal{S}^f, \forall t \in T_j^f \quad (5)$$

$$wf_{j,t+1}^f - wf_{j,t}^f \geq 0 \quad \forall f \in \mathcal{F}, \forall j \in \mathcal{S}^f, \forall t \in T_j^f \quad (6)$$

$$wi_{j,t}^f - wf_{j,t}^f \geq 0 \quad \forall f \in \mathcal{F}, \forall j \in \mathcal{S}^f, \forall t \in T_j^f \quad (7)$$

$$wi_{j,\bar{T}_j^f}^f = wf_{j,\bar{T}_j^f}^f \quad \forall f \in \mathcal{F}, \forall j \in \mathcal{S}^f \quad (8)$$

$$wf_{j,\bar{T}_j^f}^f \leq \sum_{j' \in \mathcal{L}_j^f} wf_{j',\bar{T}_{j'}^f}^f \quad \forall f \in \mathcal{F}, \forall j \in \mathcal{S}^f : j \neq dest_f \quad (9)$$

$$wf_{j,\bar{T}_j^f}^f \leq \sum_{j' \in \mathcal{P}_j^f} wf_{j',\bar{T}_{j'}^f}^f \quad \forall f \in \mathcal{F}, \forall j \in \mathcal{S}^f : j \neq orig_f \quad (10)$$

$$\sum_{j' \in \mathcal{L}_j^f} wf_{j',\bar{T}_{j'}^f}^f \leq 1 \quad \forall f \in \mathcal{F}, \forall j \in \mathcal{S}^f : j \neq dest_f \quad (11)$$

$$\sum_{j' \in \mathcal{P}_j^f} wf_{j',\bar{T}_{j'}^f}^f \leq 1 \quad \forall f \in \mathcal{F}, \forall j \in \mathcal{S}^f : j \neq orig_f \quad (12)$$

$$wi_{j,t}^f \leq \sum_{j' \in \mathcal{P}_j^f} wi_{j',t-l_{fjj'}}^f \quad \forall f \in \mathcal{F}$$

$$\forall j \in \mathcal{S}^f, \forall t \in T_j^f : j \neq orig_f \quad (13)$$

$$wf_{j,t}^f \geq \sum_{j' \in \mathcal{L}_j^f} wf_{j',t+l_{fjj'}}^f - (1 - wf_{j,\bar{T}_j^f}^f)$$

$$\forall f \in \mathcal{F}, \forall j \in \mathcal{S}^f, \forall t \in T_j^f : j \neq dest_f \quad (14)$$

$$wi_{orig_{f'},t+s_f}^{f'} \leq wf_{dest_f,t}^f \quad \forall (f, f') \in \mathcal{C}, \forall t \in T_{dest_f}^f : t + s_f \in T_{orig_{f'}}^{f'} \quad (15)$$

$$wi_{orig_f,t}^f \leq wf_{dest_f,t+end_f-1}^f \quad \forall f \in \mathcal{F}, \forall t \in T_{orig_f}^f : t + end_f - 1 \in T_{dest_f}^f \quad (16)$$

$$wf_{j,t+\text{MINTW}-1}^f \leq wi_{j,t}^f \quad \forall f \in \mathcal{F}, \forall j \in \mathcal{S}^f, \forall t \in T_j^f \quad (17)$$

$$wi_{j,t-\text{MAXTW}+1}^f \leq wf_{j,t}^f \quad \forall f \in \mathcal{F}, \forall j \in \mathcal{S}^f, \forall t \in T_j^f \quad (18)$$

$$wi_{j,t}^f \in \{0, 1\} \quad \forall f \in \mathcal{F}, \forall j \in \mathcal{S}^f, \forall t \in T_j^f \quad (19)$$

$$wf_{j,t}^f \in \{0, 1\} \quad \forall f \in \mathcal{F}, \forall j \in \mathcal{S}^f, \forall t \in T_j^f \quad (20)$$

Constraints (1) make sure that the decision variable $co_{j,h}^f$ represents the entrance of flight f in sector j at capacity period h , with its value being 1 in case of entrance, 0 otherwise, as required by the definition of that decision variable. Constraints (2), (3) and (4) define respectively the departure, arrival and sector capacity limits, ensuring that the number of flights which may depart from (or arrive at) airport k , or enter sector j at capacity period h will not exceed the departure, arrival or sector capacity for the given capacity period. Constraints (5), (6) and (7) define the time connectivity of the decision variables, with the decision variables $wi_{j,t}^f$ and $wf_{j,t}^f$ being monotonic increasing (which is specified by the first two constraints), while $wi_{j,t}^f$ must always be greater or equal than $wf_{j,t}^f$ as the first decision variable represents the first instant of possible entering of flight f in sector (or airport) j , while the latter represents the last instant of possible entering of flight f in sector/airport j . Constraints (8) ensure the fact that if a time window for some flight f and some sector j ever opens, then it will also have to close by the last time period of definition. Constraints (9) and (10) represent the continuity of a flight. The flight path, from the airport of origin to the airport of destination is given by a sequence of sectors, which are contiguous one to the other. Constraints (9) ensure the continuity of a flight from a sector j to some following sector $j' \in \mathcal{L}_j^f$, while constraints (10) ensure that a flight f can reach some sector j if and only if one of its preceding sectors $j' \in \mathcal{P}_j^f$ was in flight f 's path. Constraints (11) and (12) ensure the uniqueness of the flight's path for each flight f , respectively by guaranteeing that a flight that has reached some sector j will reach only one of its following sectors $j' \in \mathcal{L}_j^f$, and making sure that a flight that has reached some sector j can come only from one of its preceding sectors $j' \in \mathcal{P}_j^f$. Constraints (13) and (14) stipulate that a flight cannot enter the next sector on its path until it has spent at least $l_{fjj'}$ time periods (the minimum possible) travelling through one of the preceding sectors on its current path. To guarantee that the turnaround between continued flights is respected, constraints (15) are implemented. Continued flights are those flights for which the aircraft of flight f will be used for a following flight f' , with s_f being the minimum amount of time needed to prepare flight f' for departure, following the landing of flight f . Constraints (16) impose that the total flight time does not exceed the maximum duration of the flight. Constraints (17) and (18) define the minimum and maximum size for a time window respectively. Finally, constraints (19) and (20) set the decision variables $wi_{j,t}^f$ and $wf_{j,t}^f$ as binary.

III. COMPUTATIONAL EXPERIENCE

Our preliminary computational experiments show that the proposed two-step approach gives novel insights on which

actions are appropriate to cope with local demand and capacity imbalances. The first step determines a set of time windows such that the overall cost of delay is minimized. As there might be different sets of time windows generating the same minimum delay, i.e., multiple optimal solutions may exist, in the second step we choose the set of optimal time windows which also maximizes the overall time window width. In such a way, we provide to airlines, ANSPs and airports the largest degree of flexibility to perform their operations under the constraint that the minimum achievable delay is kept constant. Equivalently, we gain information on the bottlenecks or critical flights of the system. For instance, if at the end of the second step the optimal width of a time window is equal to its minimum available value any disruption that may cause the flight not to meet this time window may produce downstream effects which lead to an increase of the overall delay.

Table I: COMPUTATIONAL RESULTS FOR THE ATFMFTW MODEL. CAPACITY EQUAL TO 11 FLIGHTS PER 15-MINUTE PERIOD

Inst.	Delay		Time Windows		TW size (%)			Comp. Time
	Cost	Time	Nr.	Max Width	5	10	15	
0	2,0	10	1515	19660	9	22	69	146
1	0,0	0	1437	18970	7	22	71	10
2	0,0	0	1625	20360	12	26	62	69
3	0,0	0	1504	19625	6	27	67	51
4	31,5	155	1436	17665	17	21	63	134
5	3,0	15	1577	20010	10	25	64	507
6	0,0	0	1409	19120	6	17	77	17
7	0,0	0	1664	21670	9	21	70	53
8	0,0	0	1590	20955	5	26	69	58
9	0,0	0	1427	19385	4	21	75	27
10	0,0	0	1585	20200	12	20	67	77
11	1,0	5	1512	20390	5	20	75	69
12	4,0	20	1659	21015	10	27	63	251
13	2,0	10	1575	20405	6	28	65	382
14	4,0	20	1666	21415	8	26	65	217
15	2,0	10	1510	19550	7	28	66	476
16	2,0	10	1680	21760	8	25	67	431
17	2,0	10	1540	20870	6	18	77	103
18	0,0	0	1548	19825	11	21	68	41
19	18,2	90	1624	20130	15	23	63	308
20	0,0	0	1616	21710	6	19	75	59
21	10,0	50	1630	20350	12	26	62	1283
22	4,0	20	1576	20420	9	23	68	137
23	0,0	0	1563	20655	8	20	72	33
24	0,0	0	1477	19785	7	18	75	17
25	7,0	35	1634	20685	8	32	61	1206
26	4,0	20	1487	19165	10	23	68	92
27	22,4	110	1635	22160	6	16	78	876
28	1,0	5	1518	19125	14	20	66	185
29	4,0	20	1634	21650	7	20	72	400

We tested our model on 30 random instances of 300 non-continued flights, 25 en-route sectors, and 5 airports (3 of which are hubs). The size of the time periods in which no delay is assigned for each flight-airport are all set equal to 15 minutes, while the minimum and maximum allowed time window sizes are all equal to 5 and 15 minutes, respectively. The time periods are 5 minutes large, and the time horizon T considered is 3 hours. According to these settings, time windows can only have three different sizes: 5, 10 or 15 minutes. Moreover, the size of a single capacity period h is defined by the number of time periods included in it. The capacity for airports (both for arrival and the departure

of flights) and sectors is initially set at 11 flights every 15 minutes. Other key parameters which are set equal to all flights are: the time to traverse a sector, the maximum delay in the time window assignment, the departure and arrival delay cost coefficients that are used to ensure fairness between flights, and the maximum extra-duration for the flight time. Then the maximum acceptable duration end_f for flight f is computed by adding the maximum extra-duration to the minimum flying time calculated between the airport of origin and the airport of destination.

Table II: NUMBER OF FLIGHTS' BREAKDOWN PER TIME WINDOW SIZE AT DESTINATION AND ARRIVAL AIRPORTS. CAPACITY EQUAL TO 11 FLIGHTS EVERY 15 MINUTES

Inst.	5:5	5:10	5:15	10:5	10:10	10:15	15:5	15:10	15:15
0	15	11	12	20	46	22	12	36	126
1	3	12	16	15	35	36	13	42	128
2	19	12	17	10	62	23	15	22	120
3	5	6	4	10	31	42	19	37	146
4	48	1	13	8	56	18	15	25	116
5	21	4	8	14	55	29	16	22	131
6	7	6	12	11	30	30	16	20	168
7	8	16	21	23	30	33	17	26	126
8	6	11	10	16	55	23	15	28	136
9	1	2	13	6	38	25	16	34	165
10	25	4	10	11	44	30	25	32	119
11	8	4	14	13	44	32	13	48	124
12	23	3	6	11	75	23	17	28	114
13	9	6	18	11	45	25	15	37	134
14	12	6	11	17	52	35	16	26	125
15	12	4	16	7	47	39	12	40	123
16	9	17	5	15	56	18	27	26	127
17	6	10	9	10	35	36	20	28	146
18	23	9	12	13	38	16	17	37	135
19	38	11	5	9	52	25	15	21	124
20	8	10	18	9	45	28	16	25	141
21	28	11	7	11	48	19	16	27	133
22	22	6	5	6	53	11	11	25	161
23	7	8	12	25	27	31	11	28	151
24	5	17	10	15	28	29	22	27	147
25	13	9	7	8	70	28	30	38	97
26	13	12	18	10	47	40	23	29	108
27	7	9	7	22	25	32	27	25	146
28	30	5	13	16	40	20	10	23	143
29	16	12	10	13	52	23	16	22	136

This airspace is modeled as a grid of squared cells. This choice allows to accommodate sectors of arbitrary shape. As the instances should represent realistic cases, airports should not be too close to each other, and they should be distributed in space. Airports are therefore randomly distributed, but the minimum distance between two airports is equal to 3 cells. Airports are also randomly subdivided between regional airports and hubs, the former only connected to hubs, while the latter connected both to regional airports and other hubs.

Flights are generated randomly choosing both the departure and destination airports; imposing, however, that there are no direct connections between regional airports. The flight path unfolds among adjacent cells.

The main results obtained so far are illustrated in Table I. The first column indicates the instance number. The second column shows the solution of the first step, i.e., the minimum cost of delay whereas the third column shows the total delay in minutes. We observe that in 40% of the cases all the flights'

requests can be accommodated as no delay is produced. When there is a delay in a few instances the average delay per flight is between 10 and 31 seconds, being lower in all other cases. Thus this example depicts a situation with moderate congestion. The forth and fifth columns respectively indicate the total number of time windows existing in the system for every instance and the associated total width (in minutes), i.e., the optimal solution of the second step. The following three columns describe the percentage size (or width) distribution of such time windows. We observe that in the instances under study a share from 20% to 40% of the time windows cannot reach the maximum size of 15 minutes. In particular, a share between 4% and 17% of time windows needs to remain at the minimum width of 5 minutes. These time windows are responsible for the final total delay, and thus represent the bottlenecks of the system. The flights associated to them are the critical flights as no slack time is available in case of unforeseen events. Finally, the last column shows the computational time (in seconds) needed to solve every instance on a Intel Core 2 Duo CPU at 3.00 GHz and with 3.23Gb of Ram.

Additional information on the degree of flexibility granted to flights is available from Table II. In this table, for each instance (rows of the table) we report the number of flights with specific time window sizes. The header of the columns represents the width of the time window at the airport of departure and arrival, respectively. For instance, column 5:5 shows the number of flights with 5-minute wide time windows both at the airport of departure and arrival. So, as an example, in instance 4, 48 flights (i.e., 16%) have both departure and destination time windows of 5 minutes. These are very constrained flights, whose operations need to be tightly executed at both airports. On the other side, a large share of flights (from 97 to 168 depending on the instance) enjoys the largest flexibility as they have both departure and arrival time windows of 15-minute size (see the last column of Table II).

The available capacity at airports and sectors obviously influences the size of the time windows and the delay assigned. To analyze such capacity effects on the system, we consider different values of sectors and airports' capacity, more precisely 11, 12, 13 and 14 flights every 15 minutes. As expected, we observe that as the capacity raises the cost of the delay decreases and the total width of the time windows increases (see Table III where the average values over the 30 instances are shown).

Table III: AVERAGE VALUES OF THE 1st AND 2nd STEP OF THE ATFMTW MODEL, AND AVERAGE COMPUTATIONAL TIME, W.R.T. CAPACITY

Capacity	Delay Cost	MAX TW Width (min)	Comp. Time (sec)
11	4,14	20289,50	257,23
12	0,43	21087,17	91,09
13	0,03	21637,17	33,94
14	0,00	22024,67	26,08

The share of 15-minute time windows monotonically increases with the capacity (see Figure 1). We also notice that in situations where there is practically no congestion (as in the

case with capacity 14 where all 30 instances have no delay), a non negligible share of minimum size time windows still exists. Thus we cannot rule out the presence of bottlenecks and critical flights even when the system is apparently not under pressure.

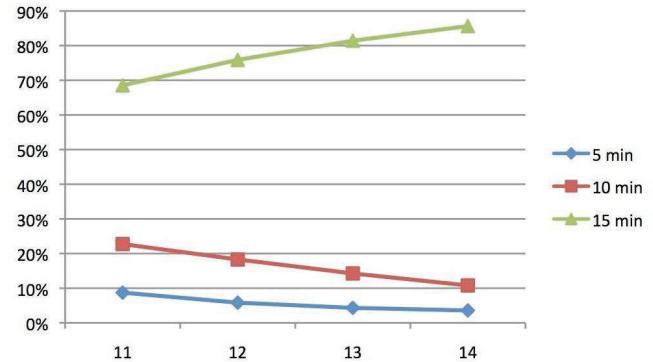


Figure 1: Time Window size w.r.t capacity

To improve the computational results that have been presented, a new version of this model is under development. This new version uses model [3] to execute the first step (delay minimization), determining the opening instants of the time windows, to set all the $w_{i,j,t}^f$ variables. This allows to reduce the size of the second problem drastically, as it will have to determine the values for the $w_{f,j,t}^f$ variables which maximize the total size of the time windows. This approach allows to obtain near optimal solutions, compared to those obtained with the original model; but the computational times are greatly reduced. Indeed, the new second step took an average computation time of 2 seconds to be solved on a set of instances for which our original model's second step needed more than 30 minutes to provide a solution. The huge computation times improvement is therefore evident, and this new approach currently looks very promising.

IV. CONCLUSIONS

This paper introduces time intervals of variable width, the so called time windows, where flights are allowed to depart, arrive or enter a sector. Each flight is expected to plan and execute its operations to comply with the sequence of its time windows (which might be of various size) from the departure to the destination airports. Analogously, the same airport or sector may have time windows of different size associated to different flights.

We present a mixed-integer programming formulation to determine the largest cumulative size of the time windows provided that the minimum total cost of delay for all flights is attained. This approach indicates to each flight the maximum available degree of flexibility to perform all its operations without providing any degradation to the system performances in terms of cost of delay. Additionally, by detecting the time windows of the smallest size (i.e., 5 minutes in our setting) it is possible to identify which flights are more constrained than

others, and which airports or sectors impose limitations on the remainder of the system.

Our preliminary results, based on small-scale random instances with moderate or low congestion, confirm the potentialities of the proposed approach. The flexibility granted to flights monotonically increases with the capacity while the system delay simultaneously decreases. We also show that apparent kind situations with no congestion may contain non negligible shares of minimum size time windows, thus indicating the existence of bottlenecks and critical flights.

In the continuation of this study we plan to further analyze how the size of time windows is distributed within the system, e.g., extending the focus on the whole sequence of time windows for a flight and not only at the departure and destination airports as in Table II. The evaluation of the spatial distribution of the time window size would also provide additional insight: airports or sectors with a large number of small time windows would be identified as critical resources or bottlenecks for the system.

An additional contribution of this work may arise by comparing the total delay costs from the ATFM-TW model and some other classical formulations of the ATFM problem (see Section I). All these latter models aim to minimize the overall cost of the delay that must be assigned to flights. To attain this goal, they identify the time period for each flight to arrive in every sector that can be flown by it, origin and destination airports included. The width of these time periods is fixed for all flights and all sectors and is usually rather large (15 min).

The mathematical model presented in this paper overcomes this limitation as it defines time windows of variable sizes for flights to execute their actions. Thus we introduce a degree of flexibility into the system that can be exploited to reduce the overall amount of delay. In fact, the time windows of fixed size become a feasible solution, but not necessarily optimal, for the ATFM-TW problem.

A new, computationally efficient version of the model is under investigation. Preliminary results show the viability of this new approach, which allows to compute second step near optimal solutions in short computational times; but a thorough computational study hasn't been carried out yet.

Finally, some modeling effort is also required to strengthen the proposed formulation and solve instances of larger scale together with more congested configurations.

ACKNOWLEDGMENTS

Luca Corolli and Lorenzo Castelli acknowledge the support from the European Commission under the project Contract-based Air Transportation System (CATS), TREN/07/FP6AE/S07.75348/036889.

REFERENCES

- [1] R.J. Hansman and A. Odoni. Air Traffic Control, in *The Global Airline Industry*, P. Belobaba, A. Odoni and C. Barnhart (eds.), 377-403, Wiley, 2009.
- [2] A. Majumdar, *Understanding En-Route Sector Capacity in Europe*, in European Air Traffic Management: principles, practice and research, AJ Cook, Ashgate Publishing Limited, Hampshire, 65-95, 2007.
- [3] D. Bertsimas, G. Lulli, A. Odoni, *The Air Traffic Flow Management Problem: An Integer Optimization Approach*, Operations Research, to appear.
- [4] EUROCONTROL, *An Assessment of Air Traffic Management in Europe during the Calendar Year 2007*, Performance Review Report, 2008.
- [5] A.R. Odoni. The Flow Management Problem in Air Traffic Control, in *Flow Control of Congested Networks*, A.R. Odoni, L. Bianco and G. Szegö (eds.), 269-288, Springer, Berlin, 1987.
- [6] M. Terrab, A.R. Odoni. *Strategic Flow Management for Air Traffic Control*, Operations Research, 41, 138-152, 1993.
- [7] P.B. Vranas, D.J. Bertsimas and A.R. Odoni. *The Multi-Airport Ground Holding Problem in Air Traffic Control*, Operations Research, 42, 249-261, 1994.
- [8] M. Helme. *Reducing air traffic delay in a space-time network*, IEEE International Conference on Systems, Man and Cybernetics, 1, 236-242, 1992.
- [9] K. Lindsay, E. Boyd, R. Burlingame. *Traffic flow management modeling with the time assignment model*, Air Traffic Control Quarterly, 1 (3), 255-276, 1993.
- [10] D. Bertsimas, S. Stock, *The Air Traffic Management Problem with Enroute Capacities.*, Operations Research 46, 406-422, 1998.
- [11] G. Lulli, A. Odoni, *The European Air Traffic Flow Management Problem*, Transportation Science, 41, 431 - 443, 2007.
- [12] A. M. Churchill, D. J. Lovell, M. O. Ball *Evaluationg a new formulation for large-scale traffic flow management* 8th USA/Europe Air Traffic Management R&D Seminar, Napa, CA, 2009.
- [13] A. Agustin, A. Alonso-Ayuso, L. F. Escudero, C. Pizzarro. *A deterministic model for air traffic flow manangement with rerouting*, 8th Innovative Research Workshop and Exibition, Brétigny-sur-Orge, France, 2009.
- [14] R. Hoffman, A. Mukherjee, T. Vossen. *Air Traffic Flow Management*, Working Paper, 2007.

Flight Profile Variations due to the Spreading Practice of Cost Index Based Flight Planning

Wilhelm Rumler, Thomas Günther, and Hartmut Fricke
 Chair of Air Transport Technology and Logistic
 Technische Universität Dresden, Germany
 wilhelm.rumler@mailbox.tu-dresden.de
 guenther@ifl.tu-dresden.de
 fricke@ifl.tu-dresden.de

Urban Weißhaar
 Department of Product Development
 Lufthansa Systems, Division Airline Operations Solutions
 Raunheim, Germany
 urban.weisshaar@lhsystems.com

Abstract—The current paper stresses the increasing relevance of the cost index (CI) based flight planning process of the Airspace Users for the Air Traffic Management (ATM) system. Based on data analysis performed together with Lufthansa Systems this paper quantifies the speed and vertical profiles in dependence on the chosen CI. Realistic CI scenarios are developed to gain a better knowledge of the ranges of flight profiles that have to be expected by the air traffic controllers. The paper shows that in cruise a range of up to Mach 0.09 for one aircraft type due to CI variations is realistic. This corresponds to about 10% speed variations. During climb and descent the range of speed can even be higher and reach values of 96 knots, corresponding to 30% speed variations. Also the vertical speed during climb and descent is influenced by the CI. Exemplary investigations of the descent profile indicate that the optimum position of the top of descent can differ up to almost 20 NM in dependence on the CI. The paper gives a detailed overview about the achieved results and briefly discusses the implications to the ATM system.

Index Terms—Cost Index, Time Costs, Fuel Management, Business Trajectory, Delay Management, Air Traffic Management, ECON Speed.

I. INTRODUCTION

In 2008 a Safety Alert due to an increased range of observed speeds in the airspace between identical aircraft types was issued by Eurocontrol [1]. With it, Aircraft Operators and Air Navigation Service Providers were invited to share their experiences regarding the appropriate reasons and consequent effects on the ATM system. Responses clearly pointed out the current lack of information regarding the cost index (CI) based flight planning process respectively the associated effects on flight profiles. The resulting uncertainties in trajectory prediction affect the provision of the Air Traffic Control services regarding increased controller workload and reduced capacity as well as potential effects on the safe separation of aircraft. In view of the modernization of the ATM system within the next decade and the aspired service-oriented approach the appropriate challenges will even rise. Airspace Users' requirements including their wish to fly close to the optimum trajectory will strongly influence the future ATM system.

The current paper stresses the high priority of CI based flight planning for airlines and contributes to an improved overall

understanding of the associated requirements for the ATM system. First a short introduction concerning the background of the CI and the appropriate cost factors as well as the optimization criteria for airlines is given. Afterwards, the effects of different CI on the flight profile are quantified based on data analysis performed together with Lufthansa Systems and its flight planning tool Lido/Flight (former Lido OC). Applied methodology includes different scenarios concerning aircraft types, flight distances and CI regimes with particular focus set to the effects on cruise speed and vertical profiles. Based on the development of cost scenarios a realistic range of speeds as well as climb and descent distances are discussed and assessed regarding the impact onto the ATM system. The paper concludes with an outlook concerning further research and development in the field of CI based flight planning.

II. BACKGROUND AND BASICS

A. Airline Operating Costs

For a comprehensive understanding of the cost index concept it is essential to have a closer look on all operating costs with effects on the in-flight performance of an aircraft. In the following, a short overview on the entire cost structure of a typical airline operator is provided in order to identify all relevant costs related to a certain flight operation and in particular those costs directly related to in-flight performance.

Basically, operating costs of an airline consist of direct and indirect costs illustrated in Figure 1 [2], [3]. Direct operating costs (DOC) cover all costs related to the flight operation of an aircraft. On the contrary indirect operating costs are independent and not connected to the operation of an aircraft mainly including expenditures for administration and distribution. Direct operating costs can be subdivided into a fixed and a variable part. Fixed direct costs are related to the operation of the aircraft but cannot be influenced by the flight event itself. These are costs for depreciation, insurance and the fixed part of maintenance and crew costs. On the contrary all variable direct operating costs are directly addressable to the flight event. Hence, a higher share of variable DOC regarding all operating costs enables an increased cost control in the frame of the flight planning process. Items that can be classified as

variable DOC are listed in Table I with their associated cost drivers¹.

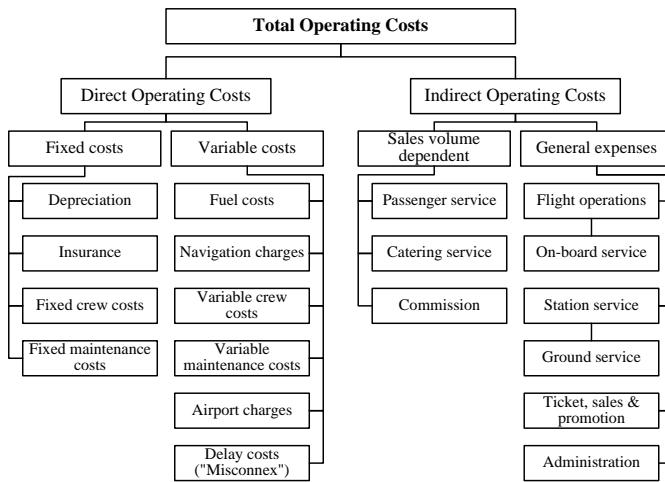


Figure 1. Airline operating costs [2], [3]

Table I
VARIABLE DIRECT OPERATING COSTS AND RELATED COST DRIVERS

Cost item	Cost driver
Fuel costs	Fuel burn
ATC charges	Airspace, distance, MTOM
Time-related maintenance costs	Flight duration
Time-related crew costs	Flight duration
Delay (Misconnex) costs	Length of delay (non-linear)

B. The Cost Index

Despite the fact that more than one-third of all airline operating expenditures are spent on fuel [4] it is obvious that optimizing a flight profile only by minimizing fuel costs is not sufficient. Economic flight planning rather takes into account all costs that are influenced by the flown trajectory and in-flight performance. To reach the most economic flight trajectory the outcome of flight track, flight profile and in-flight performance has to minimize the sum of all cost positions shown in Table I. Therefore, modern flight planning tools like Lido/Flight use algorithm for both lateral and vertical trajectory optimization. Whilst a cost function for lateral optimization has to consider all cost items listed in Table I, the cost function for vertical optimization takes into account all costs except Air Traffic Control (ATC) charges as they are neither time- nor fuel-related. Generally, ATC charges depend on the airspace charging system of the Regional Enroute Agency, the unit rate of the Air Navigation Service Provider and the flown distance as well as the Maximum Take-off Mass (MTOM) [5]. Consequently, the most economic in-flight performance and

¹Service charges for ground handling processes are also part of variable direct operating costs but cannot be optimized within the flight planning process.

resulting vertical trajectory is only based on fuel costs and the costs of time (time-related maintenance costs, time-related crew costs and delay costs).

Defined as the ratio between time-related costs and costs of fuel the cost index estimates the worth of time in relation to the price that has to be paid for fuel. Basically, the cost index expresses the time costs with the fuel unit being the currency. The CI is defined by the following formula [6]:

$$CI = \frac{C_t}{C_f} \quad (1)$$

with

$$\begin{aligned} CI &\text{ cost index} & [kg/min]_{\text{Airbus}} & [100lb/h]_{\text{Boeing}} \\ C_t &\text{ specific time costs} \\ C_f &\text{ fuel price} \end{aligned}$$

Since a large number of Flight Management System (FMS) vendors have been established in the market, two different units for the CI are generally used depending on the specific aircraft type. Airbus uses CI values with unit kg/min whilst Boeing defines the CI with 100lb/h (corresponding to 0.756 kg/min). Looking into detail at the CI equation the range of feasible cost indices can be identified. In case of nonexistent time costs the minimum CI of 0 is applied. If, in contrast, time costs are extremely high and/or the price of fuel negligible, very high values can be achieved. For instance, assuming time costs of 15 EUR/min for the A320 and a fuel price of about 0.45 EUR/kg a CI of 33 kg/min represents the optimum. The upper limitation of the CI range depends on the particular FMS. Maximum limitations are 999 kg/min for most Airbus aircrafts and 9999 100lb/h in case of the Boeing 747.

The CI is the key input value for the calculation of the speed and the vertical trajectory based on the most economical in-flight performance. Generally it is given to the pilot within the briefing package provided by the dispatch and entered into the FMS as part of the flight preparation. The flight profile calculation is done by the aircraft's integrated FMS. The calculation process is performed immediately before block-off time and, if required, during the flight in order to adapt the in-flight performance when conditions are changing. Consequently, a trajectory is calculated that balances the costs of time and fuel in order to minimize the sum of all direct operational costs.

C. ECON Speed

The crucial parameter when balancing time-related and fuel costs is the speed. Depending on the entered CI the Flight Management Computer (FMC) calculates the most economic (ECON) speed for every phase of the flight. For the minimum CI boundary of 0 time costs are neglected and fuel costs are reduced to minimum. In this case the ECON speed will equal Maximum Range Cruise (MRC) speed. For a high CI the ECON speed increases in order to reduce time costs and a speed up to the operational limitation of the aircraft is possible.

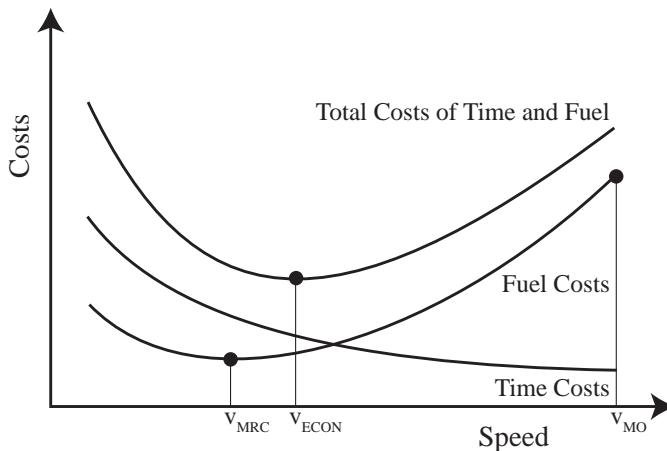


Figure 2. Impact of speed on operating costs

Figure 2 presents both fuel and time-related costs depending on the speed.

According to this figure flying with speed below or above to the ECON speed corresponding to the optimum CI will cause an increase of total costs. For low speed the fuel savings will not compensate the inevitable higher time costs and vice versa for higher speed. It must be stated that on closer examination the relation between speed and operating costs is more complex because the ECON speed varies significantly in dependence on the flight conditions. Fuel cost curve is dependent on the gross weight of the aircraft, assigned flight level (FL), the air temperature as well as the wind conditions. Time cost curve is dependent on ground speed and with it strongly influenced by the wind. This leads to shifting curves in dependence of the mentioned parameters and with it variations of the ECON speed. However, since during the flight planning process as well as the communication between the pilots and air traffic controllers normally indicated air speed (IAS) or Mach number is used, the upcoming results in Section IV refer to these speeds as well (by default with no wind and no deviation from ISA conditions).

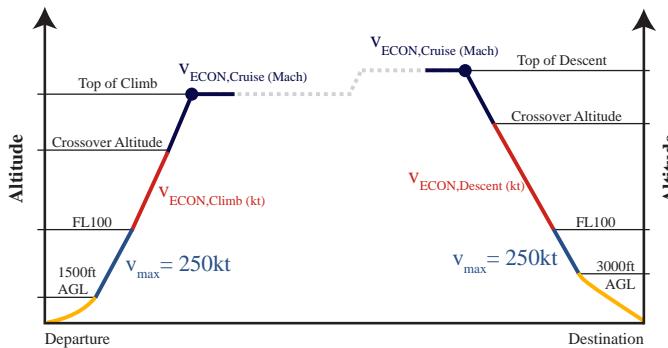


Figure 3. Speed during climb, cruise and descent, based on Lido/Flight specification for ECON Climb/Descent [7]

As depicted in Figure 3 three different speed regimes can be identified for each flight. The speed below FL100 is restricted

to 250 kt IAS according to ATC rules. As the ECON speed of most aircraft types is equal or higher for this altitude IAS is 250 kt below FL100 for almost every flight. Between FL100 and the crossover altitude ECON climb speed is measured in knots (IAS), above crossover altitude ECON Mach is applied during the remaining climb and cruise. The ECON speed is corrected during cruise phase with every step climb and when mass reduces due to fuel consumption. When reaching the top of descent same procedures are applied vice versa.

III. METHODOLOGY

In order to analyze all effects of the cost index concept and corresponding ECON speeds a simulation was undertaken using the state-of-the-art flight planning tool Lido/Flight. More than 4,000 operational flight plans have been calculated considering different aircraft types and realistic boundary conditions. In the following a short overview of the simulations setup is provided. Furthermore, in order to enable an assessment of flight profiles based on realistic cost indices, the development of appropriated cost scenarios is presented.

A. Simulation Setup

1) Aircraft Types: Every type of aircraft has its own characteristics and is therefore individual in size, in-flight performance, flight efficiency and popularity. For a representative but also feasible survey a limited sample with three different aircraft types was chosen covering national short-haul flights as well as international long-haul flights.

With the Airbus 320, Airbus 330 and the Boeing 747 three aircraft types were selected that are each among the top ten of the popularity ranking in Europe [8]. It can be assumed that with regard to size and in-flight performance, aircrafts of the same family or aircrafts of competing manufacturer will show similar results. For a validation of this assumption it is strongly recommended to continue with further studies using the same approach.

Table II
AIRCRAFT TYPES OF THE SAMPLE

Aircraft	Airbus 320-214	Airbus 330-323	Boeing 747-438ER
Engine (#)	CFM56-5B4/P (2)	PW4168B (2)	CF6-80C2B5F (4)
DOM [t]	44	122	185
MZFM [t]	61	173	252
MLM [t]	65	185	296
MTOM [t]	77	230	413
Fuel capacity [l]	23859	97530	241140
Passenger capacity [PAX]	150	295-335	416-524
Payload ² [kg]	11764	39167	55018
Max altitude [FL]	398	410	450
Max ECON speed	340 kt/M0.80	330 kt/M0.86	364 kt/M0.92
Cost index range	0-999 kg/min	0-999 kg/min	0-9999 100lb/h
Range [NM]	3000	5650	7670

²The payload was defined by taking the average load factor of the relevant distance classes of the database from the Association of European Airlines (A320: 69.2%, A330: 77.1%, B747: 81.9%) [9]

2) *City Pairs:* For a sample of applicable city pairs the study was focused on connections with an assumed high potential for delay costs. Especially flights to airports with many interconnecting flights (Hubs) can be seen as very critical regarding probable missed connections (Misconnex). If additionally the turnaround processes are planned with a very tight schedule a modification of the CI likely happens in order to reduce flight time and thus delay costs.

The airport of Frankfurt/Main (EDDF) is the third biggest airport of Europe after London-Heathrow and Charles de Gaulle in Paris. The remarkable rate of transfer passengers with over 50% makes Frankfurt to Europeans airport with the highest interconnection rate (London 35%, Paris 32%)³. Located in a central area of Europe it is the main hub of Lufthansa. Consequently, it was chosen as the destination airport for all city pairs of the sample in this study.

The selection of the departure airports was based on the 2009 flight schedule of Lufthansa. Due to the limited scope of the study only a feasible number of departure airports were used. They were selected in such a way that a homogenous distribution in geographical location (great circle distance to Frankfurt) and frequency could be achieved. Finally 30 city pairs were defined, each for the Airbus 320 and the Airbus 330. In case of the Boeing 747 two additional city pairs were added due to the extended range capability. In total 80 different departure airports with a distance range from 162 NM (EDDL) to 6449 NM (SAEZ) were applied (12 of these served by both the A330 and the B747).

3) *Flight track:* For each city pair an identical routing was applied for the entire cost index range. The routing was optimized using a Minimum Cost Track with the default value for time costs and the actual fuel price of the Lido/Flight database.

4) *Fuel Policy:* The fuel policy was based on the Commission Regulation (EC) No 859/2008 (EU-OPS) 1.255⁴. Alternate fuel was simulated with a fixed amount of fuel to reach safely the alternate airport Frankfurt Hahn (EDFH).

5) *Weather:* To ensure a high degree of comparability all flights have been calculated under ISA conditions without any wind component.

B. Realistic ranges of cost indices

In order to enable the assessment of realistic ranges of speed and vertical profiles, first realistic ranges of future cost indices are presented based on cost scenarios. Costs caused by a flight are hard to predict as they are directly connected to the specific flight event. Whereas fuel cost can comparatively easily be assessed, especially the quantification of time cost is more difficult. This is mainly due to delay costs, which can have a significant impact on the total operating costs of the flight. Delay costs are affected by the length of delay, the number and status of all involved passengers and "network effects" on other

³http://www.ausbau.fraport.com/cms/default/rubrik/6/6963.basic_facts.htm (25/04/2009)

⁴<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:254:0001:0238:EN:PDF> (13/05/2009)

connected flights. Basically, a flight delay can cause "hard" costs for compensation such as meal or drink vouchers, the rebooking of passengers or, if a rebooking is impossible the same day, a hotel accommodation. "Soft" costs are the result of a loss in revenue due to unsatisfied passengers who abandon the airline service in future as a result of the delayed flight [10]. The amount of compensation payments is regulated by law. In Europe it is based on the Regulation (EC) No 261/2004 of the EU Parliament⁵.

A comprehensive approach of estimating delay costs was undertaken by the Westminster University of London, commissioned by the Eurocontrol Performance Review Unit. Since this study proved the consideration of an extended amount of cost factors compared to previous studies (e.g. undertaken by the Institut du Transport Aérien, published in November 2000) the current paper completely refers to the results of the Westminster study. For passenger delay costs a low, base and high scenario was defined taking into account a rise of compensation costs with increasing delay minutes. The resulting average costs per passenger and delay minute for all three scenarios are listed in Table III.

Table III
COMPENSATION COSTS PER PASSENGER AND MINUTE [10]

Delay [min]	1– 15	16– 30	31– 45	46– 60	61– 75	76– 90	91– 119	120– 179	180– 239	240– 299	> 300
Low	0.06	0.17	0.26	0.35	0.42	0.47	0.58	0.75	0.89	0.92	1.15
Base	0.13	0.36	0.63	0.89	1.11	1.24	1.47	1.75	1.98	2.03	2.40
High	0.15	0.43	0.72	1.03	1.27	1.42	1.69	2.03	2.31	2.38	2.82

Values in EUR/min

Considering the passenger capacity and average payload of the aircraft, the costs of Table III can be transformed into costs per flight and delay minute. This approach is depicted in Table IV for all three types of the study (base scenario, load factor 0.75). The coefficient for payload and passenger conversion was derived from the Westminster University approach. It has to be clarified that this calculation is not taking into account the "network effect" and higher crew and maintenance costs.

Table IV
COMPENSATION COSTS PER FLIGHT AND MINUTE BASED ON WESTMINSTER UNIVERSITY APPROACH [10]

Delay [min]	1– 15	16– 30	31– 45	46– 60	61– 75	76– 90	91– 119	120– 179	180– 239	240– 299	> 300
A320	15	42	73	103	128	143	170	202	228	234	277
A330	29	80	141	199	248	277	328	390	442	453	535
B747	41	114	199	281	350	391	464	552	624	640	757

Values in EUR/min

It becomes clear that time costs per minute increase significantly with every delay minute, in particular caused by the

⁵<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2004:046:0001:0007:EN:PDF> (31/01/2010)

increased probability of rebooking or even hotel accommodations. If the individual flight is early compared to schedule, time costs often appear to be very small due to the possibility to use up the strategic buffer typically foreseen by the network planning of the airlines. The amount of time costs is then strongly influenced by the way the crew salaries are calculated by the airlines. Since previous studies (e.g. [11]) mainly focus on delay costs only, further research is needed concerning the calculation of true time costs for early flights. Hence, within this study a minimum of 0EUR per minute is assumed. However, true time costs even for early flights are expected to be higher. Current fuel price is estimated by the IATA⁶ to 0.45 EUR per kg. Since increasing prices are rather expected for the coming years the range of fuel costs is assumed between 0.30 EUR per kg (fuel price in 2004) and 1.50 EUR per kg (peak value so far in 2008 plus an allowance of 50%). Based on the presented values Table V presents the appropriate range of cost indices that are supposed to be realistic in the future.

Table V
COST INDEX RANGE FOR A320 [KG/MIN]

	0 EUR/min (no delay)	15 EUR/min (<15 min delay)	277 EUR/min (>300 min delay)
0.30 EUR/kg (low scenario)	CI = 0	CI = 50	CI ≈ 900 (max)
0.45 EUR/kg (base scenario)	CI = 0	CI = 33	CI ≈ 600
1.50 EUR/kg (high scenario)	CI = 0	CI = 10	CI ≈ 180

The range of cost indices to be expected for the A320 includes values between 0 and 900 kg/min and as such covers almost the complete CI range of the FMS. Same analysis performed for the A330 even leads to a maximum CI of approx. 1800 kg/min that even exceeds the possible FMS range. Maximum CI for Boeing 747 accounts for 3300 given in 100lb/h (corresponding to 2500 kg/min). The following section presents the respective flight profile ranges, again with particular focus on the A320 example.

IV. RESULTS

The speed is the main control variable when changing the cost index setting. It plays the key role in terms of balancing flight time and fuel consumption. However, as described above, a different speed can have significant effects on many other flight parameters of the entire flight profile, such as the optimum altitude or the optimum point for the top of climb and top of descent. In this paper the focus is set to speed and the top of descent since these values are considered to have the strongest impact on the Air Traffic Management and Control system.

A. Range of speeds

Figure 4 gives an overview of the typical ECON speeds for the A320 during the cruise phase. The ECON speed depends

primarily on the actual altitude and mass (ISA deviation and wind components were disregarded, see Section III-A for details) of the aircraft. It will alter with the increasing flight time due to performed step climbs and the reduction of the remaining fuel amount which reduces the aircraft mass. The depicted values represent the respective average speed that was calculated by taking into account the minimum and maximum speeds between top of climb and top of descent. Thus, the average values imply the speed variation of all factors mentioned above. Each grayscaled bar represents the speed average of a distance range.

The chart of Figure 4 shows a minimum cruise speed of Mach 0.59 for a selected CI of 0 and a maximum speed of Mach 0.80 for CI 130 which results in a total deviation up to 0.21 between both CI settings. However, on closer examination the ECON cruise speed is influenced by the flown distance, strongly below 250 NM. Flights inside this distance range show a cruise speed far below than all other distance ranges. This is due to the lower cruise altitude (see Figure 5). Considering all flights with a distance of more than 500 NM a minimum speed of 0.72 is applied for a CI of 0. An ECON speed of Mach 0.80 is achieved for cost indices 100 and 130. This is equal to the defined maximum operating speed within the ECON speed range.

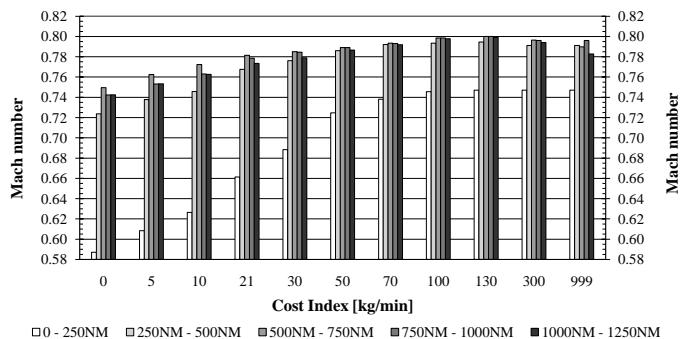


Figure 4. ECON Cruise Speed A320

It becomes obvious that the maximum speed of Mach 0.8 is reached far below the cost index limitation of 999 kg/min. Furthermore for higher CI than 130 only a very slight reduction of air speed can be noticed. This behaviour becomes clear with the analysis of the corresponding altitudes. Figure 5 illustrates the optimum flight level for each cost index of the study. Starting from a cost index around 100 the flight level decreases with higher cost indices finally ending between FL240 and FL280 for the CI limitation of 999 kg/min. This is due to the fact that ECON speed is geared to the ground speed. Because the sound of speed decreases in lower altitude the aircraft can obtain a higher ground speed if a mach number close to the maximum value of Mach 0.8 is applied in a lower flight level.

Moreover the low altitude for flights with less than 250 NM depicted in Figure 5 causes the low air speed for short distance flights that has been explained above. The altitude level of around FL260 in average is applied for this distance range.

⁶http://www.iata.org/whatwedo/economics/fuel_monitor/price_analysis.htm (31/01/2010)

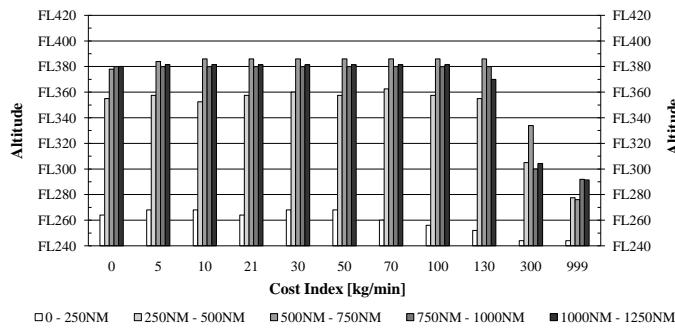


Figure 5. Average Flight Level A320

However, the identical behaviour appears for high cost indices with a reduction of the optimum altitude to FL240.

During the climb and descent phase the ECON speed above the crossover altitude (see Section II-C) is based on the speed (Mach) at the top of climb and the top of descent respectively. However, the speed below the crossover altitude seems to be the crucial parameter of both with a considerable impact on the aircraft guidance by the Air Traffic Control close to or inside the Terminal Manoeuvring Area (TMA).

According to Figure 6, the ECON climb speed for the A320 shows a similar behaviour as already presented for the cruise speed. The climb speed increases from 279 kt for the cost index of 0 up to the speed of 340 kt for cost indices higher than 130. The speed of 340 kt represents the maximum operating speed within the ECON speed range. The deviation leads to a range of 61 kt between aircrafts operating with the two mentioned cost index settings.

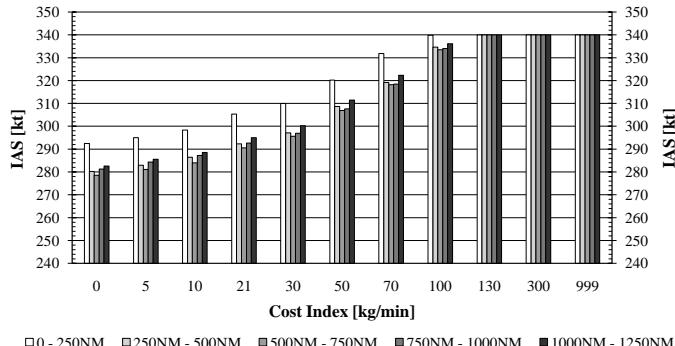


Figure 6. ECON Climb Speed A320

On closer examination of the ECON descent speeds shown in Figure 7 the speed deviation between minimum and maximum speed is considerably higher. The maximum speed of 340 kt is already applied for a CI of 100 and in case of a flight distance of more than 500 NM even for a CI of 70. Furthermore a CI of 5 and less generates an ECON speed of 250 kt. The measured descent speed range of 90 kt between a CI setting of 5 to 100 is remarkable and emphasizes the issues addressed by the current investigation.

ECON climb and descent speed is only slightly affected by the flight distance. The ECON speed of flights with a

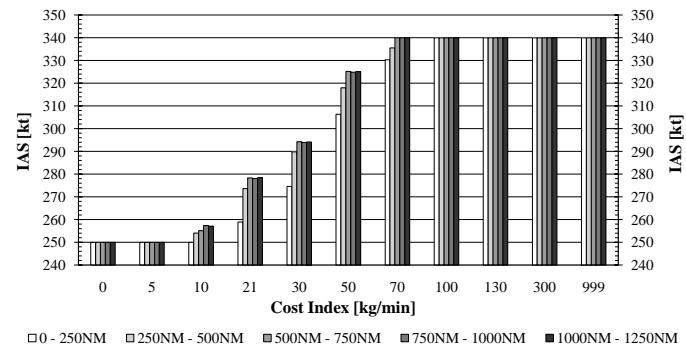


Figure 7. ECON Descent Speed A320

distance above 250 NM doesn't vary more than ± 5 kt for both the ECON climb and descent speed. It is noticeable that the ECON speed of the distance range between 0 and 250 NM is high in climb phase and lower in descent phase compared with the speed for flights with more than 250 NM. This is due to the reduced optimum flight level described above and the fact that the indicated air speed decreases relative to the ground speed for a higher altitude.

In Table VI an overview of the respective speeds of all three aircraft types is provided. Since Section III-B proved that realistic cost indices can cover almost the complete CI range supported by the FMS, speed range is presented between CI 0 and CI 999 [kg/min]. Due to impact of the reduced flight level of the A320 profile below 250 NM the speed values of this distance range are excluded. If compared to the ECON speeds of the A320, the A330 and the B747 show a similar characteristic. However, whilst the speed range between minimum and maximum speeds in climb and cruise differ less significantly the speed range in the descent phase is up to 96 kt for the B747. This is remarkable since it is equal to a variation of 30% and more than two times higher than in climb phase.

Table VI
ECON SPEED RANGE

Aircraft	A320	A330	B747
ECON cruise speed min	Mach 0.72	Mach 0.79	Mach 0.79
ECON cruise speed max	Mach 0.80	Mach 0.84	Mach 0.88
ECON cruise speed range	Mach 0.08	Mach 0.05	Mach 0.09
ECON climb speed min	279 kt	293 kt	323 kt
ECON climb speed max	340 kt	320 kt	362 kt
ECON climb speed range	61 kt	27 kt	39 kt
ECON descent speed min	250 kt	270 kt	260 kt
ECON descent speed max	340 kt	320 kt	356 kt
ECON descent speed range	90 kt	50 kt	96 kt

B. Top of descent

In addition to the speed, a crucial parameter regarding aircraft guiding and control can be identified with the top of climb and descent. Especially the optimum top of descent (TOD) is very hardly to maintain. Restrictions in speed and altitude

given by the local Air Traffic Control make a continuous idle descent difficult to achieve. In most cases this leads to an early start of descent far away from optimum and will finally result in higher fuel consumption. The TOD is optimal if the destination airport is reached by performing an idle descent with the respective ECON descent speed. Consequently, the optimum TODs for different cost indices will differ due to the wide range of ECON speeds presented in last section. For the A320 a selection of TODs associated to different cost indices are depicted in Figure 8 with the related distance to destination airport EDDF. The shown values are based on the CI specific average TOD of all flights of the sample.

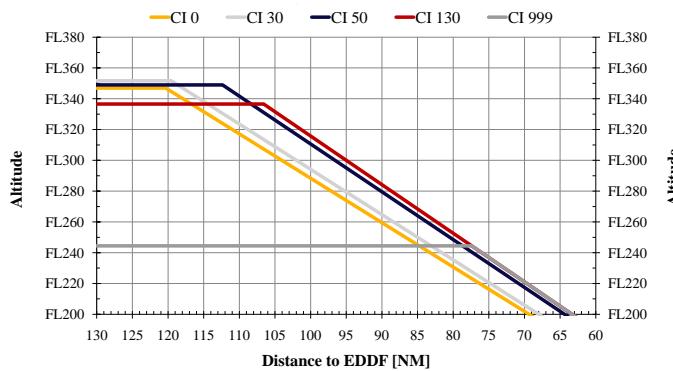


Figure 8. Top of Descent A320

With increasing cost indices and higher ECON speed the TOD "moves" towards the destination airport. If the TODs are compared around FL350 a deviation of ca. 11 NM can be stated between the minimum and maximum remaining distance to the destination airport EDDF. According to Table VII the deviation can rise to 14 NM and 18 NM in case of the B747 and the A330.

Table VII
TOP OF DESCENT RANGE

Aircraft	A320	A330	B747
Typical Altitude	FL350	FL400	FL360
Max descent length	120 NM	159 NM	138 NM
Min descent length	109 NM	141 NM	124 NM
Deviation of descent length	11 NM	18 NM	14 NM

V. IMPACT ONTO THE AIR TRAFFIC MANAGEMENT SYSTEM

Both the identified range of speeds and vertical flight profiles implies additional work for air traffic controllers due to the increasing amount of heterogeneous traffic. Although CI based flight planning has long time been a feature of airline operations, the increasing impact onto the ATM system is nowadays caused by the rising number of flights reverting to this functionality. Especially short haul flight operators consider the use of CI more and more as beneficial in view of increasing fuel prices. The widespread use of CI based flight planning software (e.g. "Lido/Flight" [12]) as well as the ongoing developments for

regional aircraft (e.g. "Pacelab™ CI OPS" [13]), stress out the appropriate changes in the flight planning processes and aircraft operations. Contrary to the past, when variations in forward and vertical speeds were mainly caused by different flight performances between the aircraft types, controllers are consequently experiencing an increasing range of speeds of the same aircraft type [1]. Two main challenges have to be stressed out in this context.

First challenge is caused by the reduced predictability of flight profiles. Prediction depending on the aircraft type is well possible based on the controller experiences, whereas the variations based on the chosen CI are dynamically influenced by the airlines and barely predictable today. Though the current ICAO flight plan includes information concerning the cruising speed and as such provides particular information, only speed changes of more than 5% have to be reported to ATC [14]. Hence, a cruising speed of Mach 0.78 indicated in the flight plan can theoretically lead to a range between Mach 0.81 and 0.75, which is adequate to a difference of approx. 40 knots in the upper airspace. Additionally, significant variations in the vertical speeds in dependence on the CI lead to uncertainties. The Single European Sky ATM Research Programme (SESAR) meets this challenge by the envisaged change of the current flight plan into a more detailed 4D Trajectory including a more detailed and precise flight profile data shared through a System Wide Information Management (SWIM) [15].

However, even if predictability can be improved, second challenge arises from the operational difficulty to manage heterogeneous traffic. Both deviations in forward and in vertical speed seem in general to influence controller workload and hence reduce capacity. Looking more into detail, current research studies estimate the influence of speed variances between a pair of aircraft on the workload as minor important than the number of vertical movements [16]. Hence it is assumed, that the avoidance of speed variances during climb and descent should have higher priority than the avoidance of speed variances in cruise. As stated in [1] aircraft operators seem to be willing to accept general speed control during climb if it is considered necessary to maintain safe separation. However, during descent the pre-advice of ATC intentions regarding time constraints is preferred due to the wish to plan the optimum flight profile under consideration of such constraints. More research on appropriate procedures in order to maximize capacity and efficiency under maintenance of an adequate safety level is required.

Consequently, the increasing range of speed due to CI based operations must both be considered as a safety issue due to the decreased potential to separate aircraft within harmonized traffic flows as well as an operational requirement that Air Navigation Service Providers have to face. As such the appropriate controller training should be conducted with particular focus on safety awareness.

VI. CONCLUSION AND FURTHER RESEARCH

Although operational flight planning is under responsibility of the airspace users, these procedures are becoming more and

more of high importance for the ATM system and the associated Air Traffic Control procedures. Hence, the understanding of the relevant principles in flight planning processes, including the CI based flight profile optimization, contributes both to the safe and efficient ATM system today as well as its convenient modernization during the next years.

The current paper proved that the variation of the cost indices with regard to the individual airline preferences lead to significantly increased ranges of flight profiles regarding speeds and vertical profiles. It is shown that in view of the expected range of time-related and fuel costs the aircraft speeds will capture almost the full range between maximum range cruise (with minimum fuel consumption per distance) and the maximum operating speed. Particular challenges arise during climb and descent where speed variations of up to 30% have to be expected without any predictability for air traffic controllers due to the missing of appropriate information given in the current ICAO flight plan format.

Situation may become more challenging in the future because the flight planning process is beginning to revert to "dynamic cost indices". While current flight plans are mainly based on an aircraft specific CI that is changed by the airlines very rarely, it is foreseen to dynamically adapt the CI to the individual conditions of each flight. This will in a first step include the cost index calculation for each individual flight during the flight planning. In a second step it will even lead to dynamic changes during the flight operations, mainly depending on changing wind conditions, weather predictions or network requirements (e.g. connecting flights) at the destination airport. This will increase the variation in speeds and vertical profiles for any particular flight independent of the aircraft operator. TU Dresden is currently developing a methodology to dynamically calculate the CI. Results will be used to assess the benefits for airlines in view of reduced fuel burn and increased punctuality as well as contributions to the environmental sustainability of air transport. With it, a more detailed understanding concerning flight profile variations to be expected in the future is aspired.

REFERENCES

- [1] Eurocontrol, *Summary of the responses to Request for Support Message "Increasing range of speeds"*, Brussels, 2008.
- [2] H. Mensen, *Handbuch der Luftfahrt*. Springer-Verlag, 2003.
- [3] J. Scheiderer, *Angewandte Flugleistung*. Springer-Verlag, 2008.
- [4] Association of European Airlines, *State of the Industry*, Brussels, 2008.
- [5] International Air Transport Association, *Airport and Air Navigation Charges Manual*, Montreal/Geneva, 2001.
- [6] Airbus, *Getting to grips with the Cost Index*, Blagnac, 1998.
- [7] V. Schmid, *Specification for ECON Climb/Descent*. Lufthansa Systems, Raunheim, 2007.
- [8] Eurocontrol, *Standard Inputs for EUROCONTROL Cost Benefit Analyses*, Brussels, 2007.
- [9] Association of European Airlines, *Traffic Trends 2007, Outlook 2008*, Brussels, 2008.
- [10] Transport Studies Group, University of Westminster, *Dynamic Cost Indexing, Technical Discussion Document 6.0, Airline costs of delayed passengers and how to estimate full network delay costs*, London, 2008.
- [11] ——, *Evaluating the true cost to airlines of one minute of airborne or ground delay*, London, 2004.
- [12] Lufthansa Systems, *The right solutions for flight operations*, Raunheim.

- [13] PACE Aerospace Engineering and Information Technology GmbH, *Pacelab™ CI OPS – The Comprehensive Solution for Cost Index Operations*, Berlin, 2009.
- [14] International Civil Aviation Organization, *Annex 2 – Rules of the Air*, Montreal, 2003.
- [15] SESAR consortium, *The ATM Target Concept*, Brussels, 2007.
- [16] J. Djokic, B. Lorenz, and H. Fricke, *ATC Complexity as workload and safety driver*. In Proceedings of the 3rd International Conference on Research in Air Transportation, Fairfax, 2008.

AUTHOR BIOGRAPHIES

Wilhelm Rumler studies traffic engineering at Technische Universität Dresden. In his study he worked on different projects regarding ATM and flight efficiency. From 2008 to 2009 he performed research on the effects of the cost index on the ATM system supervised by TU Dresden and Lufthansa Systems. Currently he is continuing with a related research in his thesis that is undertaken in cooperation with Lufthansa German Airlines in the context of the dynamic cost index concept.

Thomas Günther works at Technische Universität Dresden, Chair of Air Transport Technology and Logistics as a scientific assistant. Within his PhD thesis he is currently working on the assessment of efficiency improvements under consideration of ATM network effects. The development and application of a proper methodology shall contribute to a better understanding of according potentials regarding lateral, vertical and speed profiles as well as queue management and surface movement aspects. Thomas studied traffic engineering at TU Dresden where he received his Diploma in 2004.

Hartmut Fricke (born in Berlin, Germany in 1967) studied Aeronautics and Astronautics at Technische Universität (TU) Berlin from 1985–1991. From 1991 to 1995 he was a research fellow in Flight Operations, Airport Planning, and ATM at TU Berlin, where he completed his doctor thesis in ATM (ATC-ATFM Interface). In 2001 he finished his Habilitation on "Integrated Collision Risk Modeling for airborne and ground based systems". This included HIL experiments with an A340 full flight simulator in co-operation with EUROCONTROL Experimental Centre (EEC). Since December 2001 he has been Head of the Institute of Logistics and Aviation, and Professor for Aviation Technologies and Logistics at TU Dresden. In 2006 he was appointed Member of the Scientific Advisory "Board of Advisors" to the Federal Minister of Transport, Building and Urban Affairs in Germany.

Urban Weißhaar is currently team leader in the product development of Lufthansa Systems, division Airline Operations Solutions. He is responsible for planning and steering of strategic projects for the development of the flight planning software Lido/Flight. He represents Lufthansa Systems at international conferences and committees, dealing with ATM and environmental issues. Urban studied physics at the University of Freiburg where he received his Bachelor degree. After this he moved to the Institute of Meteorology and Climate at University of Karlsruhe where he finished his Diploma of Meteorology.

Stochastic Integer Programming Models for Ground Delay Programs with Weather Uncertainty

Charles N. Glover

Applied Mathematics and Scientific Computation Program
& Institute for Systems Research
University of Maryland
College Park, MD 20742
Cnglover@math.umd.edu

Michael O. Ball

Robert H. Smith School of Business
& Institute for Systems Research
University of Maryland
College Park, MD 20742
Mball@rhsmith.umd.edu

Abstract— Convective weather is a major contributor to air traffic delays. There is much uncertainty associated with weather predictions so stochastic models are necessary to effectively assign ground delay and route adjustments to flights. We describe a two-stage stochastic integer program for this problem. We then compare the results of this formulation to algorithms already in the literature.

Keywords: *air traffic flow management, integer programming, stochastic programming*

I. INTRODUCTION

A major priority of the Federal Aviation Administration (FAA) is to reduce airport congestion. Air traffic flow management specialists within the FAA seek to address this issue by resolving instances in the National Airspace System (NAS) where the anticipated demand exceeds airport capacity. A limitation on the number of airports that can be built, paired with a significant increase in air traffic leads to airport congestion being a primary concern. Added to the simple increase in air traffic is the more complex situation that convective weather has on causing demand to exceed capacity at airports. When the FAA predicts this to occur at an airport, they are placed into a situation where airport landing slots become a scarce resource, and must be allocated to flights through some traffic flow management initiative. One of the most advanced such procedures is that of a ground delay program (GDP). Rather than assigning delay to flights in the air, a GDP is a preemptive measure that holds aircraft on the ground before they depart their origin airports. The net effect of this is that the more costly and more risky usage of airborne delay is reduced and transferred to the ground where it is more easily managed. This also reduces the stress on air traffic managers, who have limited options once the aircraft are airborne.

GDPs were first implemented after the airline strike of 1981 [6]. Since then, they have become a growing part of our airline industry. In 2006, there were 1305 GDPs implemented in the United States [5]. The cost of these delays to the airlines and passengers is billions of dollars per year. So it is only logical that we would like for this delay to be at a minimum.

There are many roadblocks to efficient minimization of delay in a GDP. One primary such roadblock is that of equity. Before the current standard of Collaborative Decision Making (CDM) was adopted, participants felt that GDPs were implemented in an inequitable manner. Airlines (correctly) felt that, in many cases, the information they provided was used by the FAA to provide a much greater benefit to their competition than to the airline providing the information. Thus, they would provide out of date or inaccurate information. This lack of equity led to inefficient solution procedures and often resulted in more system delay. CDM was initiated to resolve these issues by instituting methods that were based on agreed upon standards and allocation procedures that provided incentives for participation with honest information [1], [9].

One of the major results of this was the ration-by-schedule (RBS) principle, which decoupled the information provided by the airlines on a day of operations and the resources they received. The basics of the RBS principle are first scheduled first served, so in a GDP the flights are kept in the order that they were originally scheduled. Some flights, though, are exempt from RBS. One set, flights that have already taken off, obviously cannot be given ground delay and must be exempt. The other set, though, is not as simple.

Because of the stochastic nature of weather, an air traffic manager is reluctant to delay a flight several hours in advance of a storm that may or may not materialize. An overly pessimistic forecast could result in some longer flights being given what, in hindsight, is unnecessary delay. To offset this, a distance radius is set from the troubled airport, and ground delays are only assigned to flights that originate inside that radius. The remaining flights are exempt from this GDP.

Ball et al. [2] developed a formal stochastic model of GDP's to gain a fundamental understanding of how giving preferential treatment to long-haul flights improves expected GDP performance. They proposed the ration-by-distance (RBD) algorithm, which allocates flights to arrival time slots using a priority scheme based on flight list ordered by decreasing flight length. This algorithm is structurally the same as RBS, but with an alternative priority scheme. They showed that RBD, under a fairly general model of GDP dynamics,

minimizes the expected delay [2]. It is easy to see, however, that RBD can generate an inequitable distribution of flight delays. To address this problem, they proposed a heuristic algorithm, E-RBD, that would seek to balance efficiency and equity. In this paper, we formulate an integer program (IP), which represents the GDP with weather uncertainty. We then show how this IP can use different objective functions to more precisely balance efficiency and equity to a larger scale than either RBS, RBD or E-RBD.

II. RELATED WORK

The GDP is a well-studied problem in aviation research. Odoni first proposed an IP model for the Ground Holding Problem [8]. Bertsimas and Stock Patterson formulated a model to address issues concerned with congestion in the National Airspace System (NAS)[4]. This model minimizes the total ground delay and airborne delay, while ensuring that the departure capacities, arrival capacities, sector capacities and time connectivity constraints are not violated. Although, the model is for the ATFM problem, it can easily be adapted to represent the Single Airport Ground Holding Problem (SAGHP) and Multiple Airport Ground Holding Problem (MAGHP). These models are deterministic and do not account for the ways that the weather uncertainty can play into the planning of a GDP.

Ball et al. studied a stochastic case of GDPs [3]. In this problem, they were concerned with Airport Arrival Rates (AARs), the number of flights the airport can receive in a given time period, in an environment where the weather is uncertain. The model takes into account an AAR distribution, and produces a planned AAR (PAAR) vector, which is the number of flights that the airport should schedule to arrive in each time period, given the stochastic nature of the weather and the probabilities of different AARs. Kotnyek and Richetta showed that a model first proposed by Richetta and Odoni could also be used to produce the PAAR vector [6]. This model is larger in size than the Ball et al. model, but because its cost function for ground delay is more general, it allows for more specific adjustments of the relationship between the costs of airborne holding and ground holding.

Both these models operate under the condition of weather uncertainty. Due to the excessive costs of airborne holding when compared to that ground holding, both papers try to avoid the situation where airport has more flights seeking landing than it has landing slots available in a given time period. But it is also possible to have a larger number of available landing slots than flights seeking landing. Such a situation can arise when an airport expects convective weather and flights are given more ground delay than necessary to offset the convective weather. In these situations the airport would like to be able to re-schedule flights to utilize this unexpected capacity. Because the papers by Ball et al. and Kotnyek and Richetta consider only the static case of stochastic ground delay programs, their models do not allow us to adjust the delays once the weather uncertainty becomes certain.

Mukherjee and Hensen presented a dynamic stochastic IP formulation for the SAGHP, which took as part of its input the possible changes the weather can take throughout the duration

of the GDP. [7] This formulation presented a scenario tree to capture all the possible changes in weather outcomes. This scenario tree can grow large in size though and can make the IP computationally inefficient.

In [2], Ball et al. consider the problem of maximizing the throughput into the airport. Here, the RBD algorithm is first proposed. The authors prove that the RBD algorithm minimizes total expected delay if the GDP cancels earlier than anticipated. In their proof, the authors were able to compare the total expected delay of the RBD allocation with that of other allocations and are able to show optimality.

III. FORMULATION

The input to the model comes from two sources: flight-based input and airport-based input. The flight-based input includes:

- The length of the flight k , $len(k)$
- The scheduled arrival time of the flight k , $a(k)$
- The arrival slot that the flight k would receive in the RBS allocation, $RBS(k)$

The airport-based input includes:

- The expected duration of the GDP
- The nominal capacity of the airport slot i , $cap_2(i)$
- The reduced capacity of the airport slot j , $cap_1(j)$
- A list of possible end times for the GDP
- A probability for each possible end time

We also base our model on some important assumptions.

- The weather has only two possible states, clear and not clear. This is done to keep the problem from growing too large and to model how GDPs are handled in practice, where a GDP is not cancelled until the weather is clear. This collapses the scenario tree and allows the problem to be modeled as a two stage stochastic IP instead of as a multi-stage stochastic IP.
- There is no layover between weather clearance time and the time the airport goes back to nominal capacity. We assume this happens immediately.
- We do not consider the airborne holding as an alternative option for ground holding. We will thus not allow for solutions that, in expectation of an early weather clearance time, send more flights to the airport at a given arrival slot than that slot will allow.
- We assume no prior knowledge of weather clearance until the weather has actually cleared.

We will formulate the IP as a stochastic IP. There are two decision stages. In a GDP, every flight must initially be assigned to a slot, and the first stage models these actions. This takes place with no future knowledge of when the weather will clear. In the second stage, a weather clearance time becomes

known and some flights can be reassigned to newly available slots.

Let $x_{k,i}$ be the binary variable which is non-zero if flight k is initially assigned to the arrival slot i . Then the following two constraint sets model the stage one restrictions. These constraints are very similar to the model proposed by Odoni [8].

$$\sum_{i \geq a(k)} x_{k,i} = 1 \text{ for each flight } k, \quad (1)$$

$$\sum_{k|a(k) \leq i} x_{k,i} \leq cap_1(i) \text{ for each arrival slot } i. \quad (2)$$

Each flight has a scheduled arrival time, $a(k)$, and the first constraint set ensures that each flight is assigned to some arrival slot after its scheduled arrival time.

Notice that we place no upper bounds on the latest arrival slot to which, a flight can be assigned. We note that deviations from a flights RBS slot, $RBS(k)$ are a measure of inequity. The E-RBD algorithm seeks to find equitable solutions by restricting how long a flight can be delayed after its RBS slot. We can make this adjustment to the model by placing this restriction, $RBS(k) + \delta$, as an upper bound on the summation in (1).

$$\sum_{i \geq a(k)}^{RBS(k)+\delta} x_{k,i} = 1 \text{ for each flight } k, \quad (1a)$$

Each slot has an initial capacity, $cap_1(i)$, the number of flights the airport can handle during the reduced capacity. So the second constraint set ensures that no slot is utilized in excess of its capacity during the GDP. The corresponding adjustment to constraint (2) is:

$$\sum_{k|a(k) \leq i \leq RBS(k)+\delta} x_{k,i} \leq cap_1(i) \text{ for each arrival slot } i. \quad (2a)$$

This completes stage one of the formulation.

In stage two we have a scenario, t , for each possible weather clearance time. Upon clearance of the weather, we assume that the number of arrival slots has immediately increased back to full capacity. In such a situation, it is very possible for flights to be assigned to earlier slots than the slot to which they were initially assigned.

The following constraint set defines a queue in each scenario of stage two amongst the slots available in that scenario. The function $earliest(k, i, t)$ maps the allocation (k, i) from stage one to the earliest arrival slot that it can be reallocated to in scenario t of stage two. If $earliest(k, i, t) = j$, then the variable $x_{k,i}$ can enter the scenario t queue at slot j , depending on whether its value is 1 or not. The variable $z_{j,t}$ is

the amount that is passed from slot $j-1$ to slot j in scenario t . The following constraint immediately follows:

$$\sum_{(k,i,t)|earliest(k,i,t)=j} x_{k,i} + z_{j-1,t} - z_{j,t} - u_{j,t} = 0 \quad (3)$$

for each arrival slot j and each scenario t

The flight dependent input to this IP is the arrival time of each flight, $a(k)$, the length of each flight, $len(k)$, and the RBS allocated slot of each flight, $RBS(k)$. Based on this input, we are able to determine $earliest(k, i, t)$ for each stage one allocation (k, i) as a pre-processing step.

An allocation (k, i) where $x_{k,i} = 1$ can be in one of three states at the beginning of a stage two scenario: It is either in the air, on the ground because its scheduled departure time has not yet passed, or on the ground because it is serving delay. The determination of which of these sets an allocation belongs to consists of checking the delayed departure time of the allocation, $i - len(k)$, and comparing it with both the weather clearance time, t , and the scheduled departure time of the flight, $a(k) - len(k)$. If the delayed departure time is after the weather clearance time, t , then the flight has departed; if it is before $a(k)$, then it is on the ground because its scheduled departure time has not yet passed; and if it is after $a(k)$ and before t , then it is on the ground because it is serving ground delay.

Depending on which state an allocation is in, there are limited recourse actions that can be taken. Flights already in the air cannot depart for an earlier arrival slot than the one which they are initially assigned. The earliest arrival slot these flights can be assigned to is thus the slot to which were initially assigned, so $earliest(k, i, t) = i$ for these flights. Flights on the ground because their scheduled departure time has not yet passed still cannot depart. In this scenario though, they will be free to depart for any arrival slot equal to or after their scheduled arrival slot. The earliest arrival slot for these flights is thus their scheduled arrival slot, so $earliest(k, i, t) = a(k)$ for these flights. Flights that are grounded because they are serving delay can depart immediately. These flights, though, cannot arrive at an arrival slot earlier than the time it takes to travel from origin to destination, so for these flights $earliest(k, i, t) = t + len(k)$.

We also need to ensure that the nominal capacity is not violated. This can be achieved by the following constraint:

$$u_{j,t} \leq cap_2(j). \quad (4)$$

for each arrival slot j and each scenario t

The objective function will consist of two components: one to measure efficiency and one to measure equity. The metric for efficiency is based on minimizing the total expected delay, which for an individual flight can be recorded as the flight's final assigned time minus its scheduled time. We do not keep track of the final assigned times for each flight in this formulation. So, instead, we can measure efficiency by subtracting the sum of the flight scheduled times from the sum of the assigned times. Because we want the expected delay, we

also multiply each term by the probability of that scenario occurring. This metric can then be represented by:

$$\sum_{t \in \text{Scenarios}} p_t \left(\sum_{j \in \text{Slots}} (j \cdot u_{j,t}) - \sum_{k \in \text{Flights}} a(k) \right). \quad (5)$$

The metric for equity will be based on the RBS solution being the most equitable solution. Each flight will then be penalized by how much later they are assigned from their RBS allocation. This can be represented as:

$$\sum_{i \in \text{Slots}} \sum_{k \in \text{Flights}} \text{cost}(k, i) \cdot x_{k,i}, \quad (6)$$

where $\text{cost}(k, i)$ is the deviation of the stage one assignment of flight k from its RBS slot, i.e. $\text{cost}(k, i) = 0$ if $i \leq \text{RBS}(k)$ and $\text{cost}(k, i) = i - \text{RBS}(k)$ otherwise.

IV. EXPERIMENTAL RESULTS

We tested this formulation using data based on GDPs run on three different dates at San Francisco International Airport (SFO) with seven possible weather clearance times. The tests were run on three different probability distributions: a uniform distribution, a distribution where the probabilities of weather

clearance were $p = \left[\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \frac{1}{64}, \frac{1}{64} \right]$ and a distribution where the probabilities of weather clearance were $p = \left[\frac{1}{64}, \frac{1}{64}, \frac{1}{32}, \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2} \right]$.

For our nominal capacity we used the Airport Arrival Rate (AAR) of 60 flights per hour, and for our reduced capacity, we used the AAR of 36 flights per hour, or more precisely, we used 6 flights every 10 minutes.

For the IP, the experiment was conducted on each day and distribution with coefficients ranging from 0 to 1, incrementing by 0.1. To obtain the E-RBD results we used an objective function that measured only efficiency and restricted the allowed assignments of a flight to only δ minutes after its RBS allocated slot, for each given δ . Both the IP and the E-RBD algorithm give the RBS and RBD solutions at their extreme parameter values. For instance, if we set the coefficient for equity to 1 in our IP, our focus is only on equity. We will then receive the most equitable solution, RBS. On the other hand, if we set the coefficient for efficiency to 1 in our IP, our focus is only on efficiency. We will then receive the solution that has the least total expected delay, which is the RBD solution.

The E-RBD algorithm has similar properties. When we set the maximum deviation, δ , equal to 0, we are not allowed to deviate from the RBS solution, which is the optimal solution in that case. If we set δ to an arbitrarily large constant then all stage one assignments are allowed, in which case, the E-RBD algorithm will output the RBD solution.

One key difference between the IP and the E-RBD algorithm, though, is the fact that the IP allows us to choose a coefficient small enough that it keeps us close to either the RBD or RBS solution, while still taking into account both equity and efficiency. For instance, the RBS and RBD solutions obtained from our IP were obtained with equity to efficiency ratios of 1 to 99 and 99 to 1 respectively. This helped find a “more equitable” RBD solution or “more efficient” RBS solution amongst the many available. We did not make such considerations with the E-RBD algorithm because the algorithm, as described in literature, is based on optimizing efficiency by placing a limitation on inequity.

SFO Day 1 Uniform Probability

Max Allowed Deviation from RBS (m)

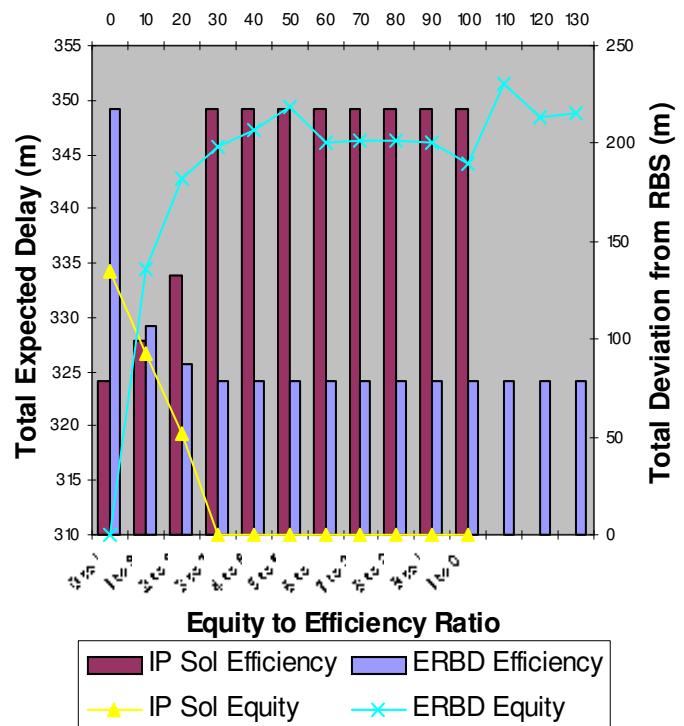


Figure 1. Equity and efficiency comparisons between the IP formulation and E-RBD solution at SFO under a uniform probability distribution on day 1.

For figures 1, 2, and 3, we have an x-axis corresponding to each solution. The top axis represents the maximum deviation we allowed for the E-RBD algorithm. The bottom axis represents the ratio of equity to efficiency we gave the objective function in the IP solution. We measure the efficiency of both algorithms with the bar graphs on left axis. The equity is measured with the line graphs on the right axis.

In this Figure 1, we see that with an equity coefficient of 0.1, we are able to obtain an IP solution, which is close in efficiency to the RBD solution, and is more equitable than any of the solutions returned by the E-RBD algorithm, with the exception of the RBS algorithm. The IP solution with an equity coefficient of 0.2 is also more efficient than the RBS solution.

Neither the IP or the E-RBD algorithm give many solutions outside of RBD and RBS. This leads to an immediate question of whether this is true in general, or just a product of this example.

SFO Day 3 Decreasing Probability

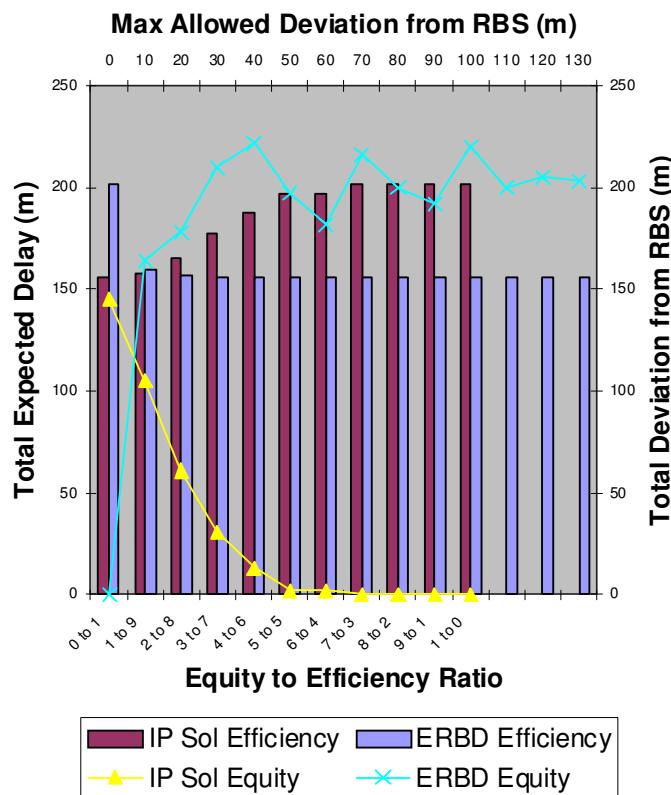


Figure 2. Equity and efficiency comparisons between the IP formulation and E-RBD solution at SFO under a decreasing probability distribution on day 3.

Figure 2 provides us with a difference between the efficiency of the RBS and RBD solutions of about 50 minutes. This was the largest difference amongst the examples we considered. Once again we notice the solution provided by the IP with an equity coefficient of 0.1 provides a solution that is efficient (here, within 3 minutes of the total expected delay of the RBD solution), and more equitable than any of the E-RBD solutions. There are also more solutions found in this example by the IP that are not RBD or RBS.

The E-RBD solution, however did not offer such a diverse set of solutions. There were two such solutions, found when δ has values of 10 and 20. Both these solutions, though, have comparable efficiency to the IP solution with an equity coefficient of 0.1. Even with such a low coefficient though, the IP solution is more equitable than both these E-RBD solutions.

Of the nine test cases considered, each one provided exactly four distinct objective function values for the E-RBD solutions – when δ ranged from 0 to 30. In each case, the solutions when $\delta \geq 30$ have an objective function value equal to the RBD solution. As stated earlier, when $\delta = 0$, this algorithm gives the RBS solution. When $\delta = 30$ we received the RBD solution in each case. So the E-RBD algorithm only provided two non-

extreme solutions to compare and contrast with those given by the IP formulation.

In light of this, we decided to run the E-RBD algorithm with smaller values of δ to see how these solutions compared to the IP solutions. The following graph compares these results to similar results given by the IP, using similar axes. Notice though the change in the δ values.

SFO Day 2 Decreasing Probability

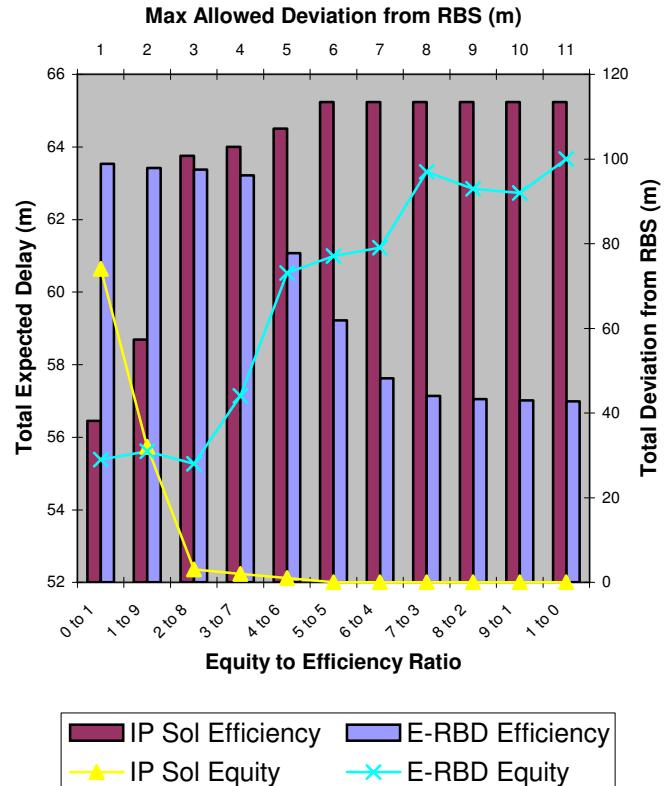


Figure 3. Equity and efficiency comparisons between the IP formulation and E-RBD solution at SFO under a decreasing probability distribution on day 2.

What we notice in Figure 3 is that the E-RBD algorithm begins to give solutions whose total expected delay savings is significant from the RBS solution at $\delta = 5$. This is at a cost of a sharp increase in the total deviation from RBS and the inequity remains high for all remaining $\delta > 5$. It is of course worth questioning if E-RBD behaves this way in general. Because E-RBD does not take total deviation from RBS into account, it is perhaps not surprising that it does not do as well in finding solutions with low values of this objective.

We were also interested in the role that the probability distribution of the weather clearance time plays on the optimal solutions to the IP. Both RBS and RBD allocate flights to slots without any regard to this probability distribution. The IP is based on a two-stage stochastic IP and thus fundamentally based on this probability distribution.

To test whether the distributions generated different optimal solutions, we used one objective function and output the objective function value of the solutions generated by each probability distribution. Because the probability distribution

does not affect the feasibility of solutions to our IP, evaluating these solutions under the same objective function will imply that different solutions with the same objective function value are both optimal. We are thus interested in where the objective function values differ. Below is a graph of the number of solutions and how it corresponds to each ratio.

Day 3 IP Solutions Comparison

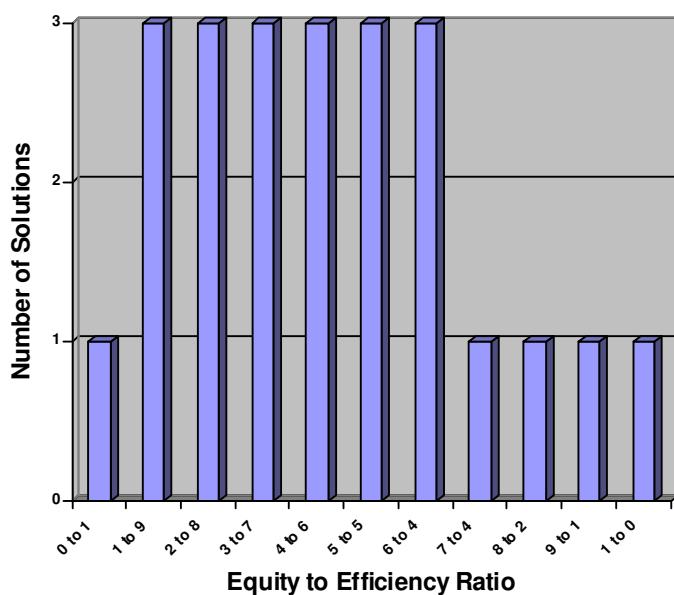


Figure 4. A count of the number of different IP Solutions under the three different probability distributions we tested.

Figure 4 shows us that when the solution to the IP is not RBS or RBD, the probability distribution can play an important factor. If we compare Figure 4 to Figure 2, we see that every instance that is not RBS or RBD, we achieve 3 different solutions for the three different distributions. In general, we found that in all examples except one, if the optimal solution was not RBS or RBD, we received at least two different solutions from the three different distributions.

V. CONCLUSIONS AND FUTURE WORK

We were able to use the IP formulation to generate solutions which have a comparable amount of total expected delay to the E-RBD solution, but which are more equitable. The formulation also gives solutions that are comparable in both equity and efficiency to other rationing principles in the literature such as RBD and RBS.

One thing we noticed here is that both the range of δ considered here and the coefficients for the equity to efficiency ratio were of a very general variety here. It would be interesting to see how these solutions compare with values inside a specified range, say $\delta \in (0, 30)$ and equity coefficients between 0.1 and 0.2. We suspect that the results for E-RBD would be of little difference because the algorithm does not take the total equity of the GDP into consideration when considering allocations. However, it would be of interest to know how close we can have our IP solution to RBD in terms

of efficiency, while still remaining within a certain deviation from the RBS solution.

In many instances, though, we also receive an IP solution, where stage one is deemed inequitable by E-RBD because too much of the delay is given to one flight or a set of flights. This means that these solutions are not feasible to the more restrictive E-RBD constraints. The only difference in constraints, though, is the limitation placed on flights by δ . This implies that the IP solution may violate this δ constraint for some flights, even when it has low total deviation from RBS.

An additional area to consider, then, is placing equity limitations on each flight in this IP, while still measuring the objective function as a linear combination of equity and efficiency. A benefit of this would be that we would be able to ensure that no single flight or set of flights receives any bulk of the delay, which was a motivating factor behind the development of the E-RBD algorithm, while at the same time minimizing total expected delay and total deviation from the RBS allocation.

We are also currently investigating other properties of this model such as its polyhedron structure and if there is a general form for the optimal solution with similar objective functions. There are also a host of other problems which have a similar structure to this problem, so gaining more understanding of this IP can prove helpful to better understanding those problems as well.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the National Aeronautics and Space Administration Airspace Systems Program under ARMD NRA: NNH06ZNH001

REFERENCES

- [1] Ball, M.O., C. Barnhart, G. Nemhauser and A. Odoni. 2007, "Air Transportation: Irregular Operations and Control", *Handbook of Operations Research and Management Science: Transportation*, C. Barnhart and G. Laporte, eds, Elsevier, Amsterdam, pp 1-67.
- [2] Ball, M.O., R. Hoffman, and A. Mukherjee. 2010 "Ration-By-Distance with Equity Guarantees: A New Approach to Ground Delay Program Planning and Control", *Transportation Science*, 1-14.
- [3] Ball, M.O., R. Hoffman, A. Odoni, and R. Rifkin. 2003. "A Stochastic Integer Program with Dual Network Structure and its Application to the Ground-Holding Problem", *Operations Research*, 51, 167-171.
- [4] Bertsimas, D., and S. Stock Patterson. 2000. "The Traffic Flow Management Rerouting Problem in Air Traffic Control: A Dynamic Network Flow Approach", *Transportation Science*, 34, 239-255.
- [5] Kotnyek, B., and O. Richetta. 2006. "Equitable Models for the Stochastic Ground-Holding Problem Under Collaborative Decision Making", *Transportation Science*, 40, 133-146.
- [6] Manley, B., Sherry, L., 2008, "The Impact of Ground Delay Program (GDP) Rationing Rules on Passenger and Airline Equity", ICRAT.
- [7] Mukherjee, A., and Hansen, M., 2007 "A Dynamic Stochastic Model for the Single Airport Ground Holding Problem", *Transportation Science*, 41, 444-456.
- [8] Odoni, A.R., 1987, "The Flow Management Problem in Air Traffic Control," in *Flow Control of Congested Networks*, Odoni, A.R., L. Bianco, and G. Szego, eds., 269-288, Springer-Verlag, Berlin.
- [9] Vossen, T., M.O. Ball, R. Hoffman, and M.Wambsganss. 2003. "A General Approach to Equity in Traffic Flow Management and its

Application to Mitigating Exemption Bias in Ground Delay Programs”,
Air Traffic Control Quarterly, 11, 277-292. (a preprint was published in
Proceedings of 5th USA/Europe Air Traffic Management R&D Seminar,
2003).

Enhanced Wind Magnitude and Bearing Prediction

Onboard algorithm

Petr Krupanský, Tomáš Neužil, Eva Gelnarová

Advanced Technology, Aerospace
Honeywell International
Brno, Czech Republic

Jiří Svoboda, Petr Mejzlík, Martin Herodes

Advanced Technology, Aerospace
Honeywell International
Brno, Czech Republic

Abstract—The article deals with the aircraft onboard wind prediction if there is no up-to-date accurate weather forecast available. The simple method for extrapolation of measured wind dynamics is presented. Also the algorithm for blending average wind trends with measured data is presented.

Keywords - wind prediction, RUC (Rapid Update Cycle) analysis, wind trends

I. INTRODUCTION

The proposed Enhanced Wind Prediction Algorithm (EWPA) is intended to be a tool for the onboard wind magnitude and bearing prediction in case of absence of a suitably accurate meteorological forecast in the Flight Management System (FMS). The further motivation for this study and for an algorithm development is the fact that studied meteorological forecasts [5] have a lower accuracy than onboard measured data in vicinity of the aircraft.

The key question which the enhanced algorithm development tries to answer is: "Is it possible to improve actual onboard predicting algorithm even without a valid up-to-date wind forecast available onboard?"

As an answer to this question the following approach has been designed and tested. Since the actual prediction algorithms deals only with wind magnitude and bearing data for a level flight, the proposed one focus also only on these quantities. However, any other variables (such as temperature, humidity, etc.) can be predicted in a similar manner.

In absence of up-to-date wind forecast onboard [6], the algorithm currently used in FMS utilizes an actual wind measured by sensors as the measurement prediction (e.g. prediction of the future wind behavior). The values of the wind measurement prediction are directly projected to all points along the (planned) flight path. This approach does not consider the dynamic change of the wind during the flight path and thus does not evaluate wind behavior dynamics. This is a serious limitation of the described approach, because the wind dynamics is an important factor, as stated in [3]. The proposed approach combines the local development of onboard measured data together with average wind trends, which characterize the wind behavior in broader horizon/more distant segments of the trajectory. The average trends of the winds are supplemented with the information about their standard deviations. The identification of the typical trends together with their description can be derived from any suitable weather

database, which collects data for sufficiently long time period. In case of the presented study, the RUC databases have been used [5]. The methodology used for the analysis of weather data and resulting wind characteristics are described in [3].

II. MEASUREMENT PREDICTION MODEL

The EWPA is initially intended to be a tool for the CRUISE phase of the flight designed with the respect to the possible embodiment of the other phases of the flight. The design was constrained also by computation performance of the FMS [1].

Prior the flight, the preprocessed average weather characteristics are loaded to the FMS. During the flight, the data about the wind magnitude and bearing are periodically measured and stored with a given frequency. The set of the last measurements is used for determination of the parameters of the simple dynamic model for the measurement prediction (prediction of the future wind evolution) in local vicinity of an aircraft.

The following text describes the steps of EWPA:

- The prediction of the wind measurement magnitude values based on on-board measured data.
- The blending of the measurement predictions with the typical weather trends along the trajectory.

The prediction of the wind bearing values is based on the same algorithms and therefore only the magnitude part of the algorithm is presented.

A. Wind Magnitude Prediction

The measurement model is used for the magnitude measurement prediction. The first order dynamic model was chosen as a suitable and simple description of the wind magnitude behavior. The nature of the wind behavior is tended to be steady than continuously changing [3]. This way of behavior is introduced by the saturation limit of the wind measurement model [7]. The function of the measurement model can be constructed as:

$$Mag_{Measure}(d) = \text{sgn}(k)K \left(1 - e^{-\frac{d_0-d}{D}} \right) + Mag_{d_0} \quad (1)$$

Where:

$Mag_{Measure}$... the value of the predicted wind magnitude,

The work was supported by Honeywell International and European Commission (ERASMUS project)

$Mag_{d_0} \dots$	the value of the wind magnitude in the point d_0 ,
$d_0 \dots$	the distance to the point of the last measurement,
$K \dots$	the absolute value of the measurement model saturation,
$k \dots$	the angular coefficient of the tangent line (eq. 3),
$D \dots$	the distance from the actual aircraft position d_0 to the point of the model saturation (eq.4),
$d \dots$	the distance.

For the determination of the measurement model parameter D the parabolic regression is fitted to the wind data. The parabolic interpolation allows simple description of the dynamics of the wind development and its computational demands are not high at the same time. The equation of the parabolic interpolation is:

$$y = ad^2 + bd + c \quad (2)$$

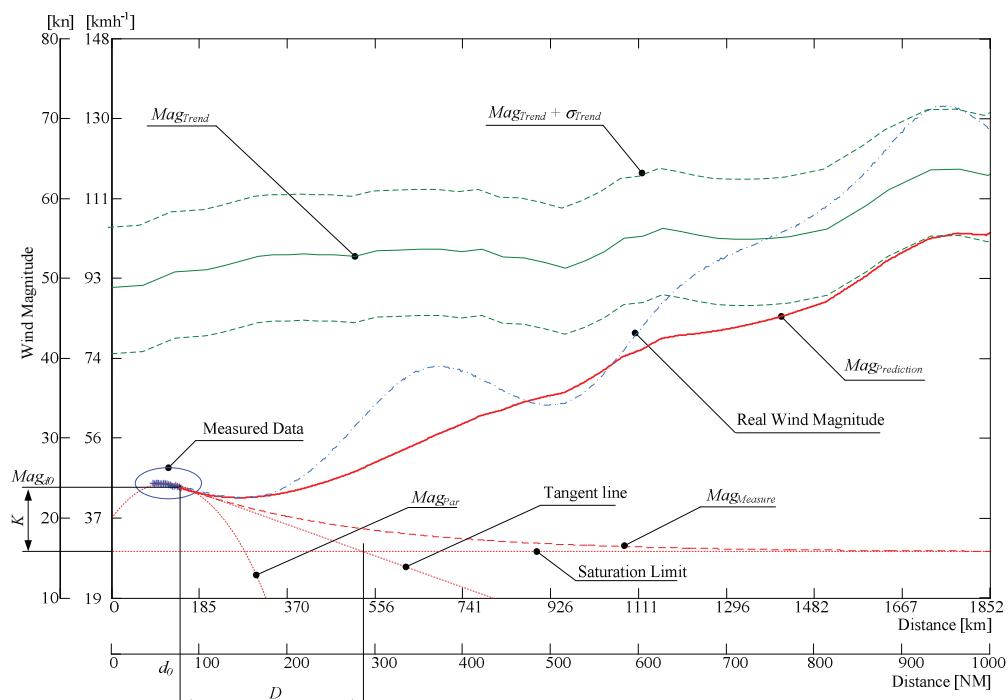


Fig.1. Wind Magnitude Prediction

B. Standard Deviations

The second part of the measurement model provides the estimation of the standard deviations associated with the predicted wind magnitudes at the trajectory points. The uncertainty of the measurement prediction increases with the increasing distance of the trajectory points from the current aircraft position d_0 , (represented by an increase of the standard

where:

- $a, b, c \dots$ the parameters of the parabolic function,
- $y \dots$ the value of the interpolated data.

In the next step, the resulting parameters of the interpolation are used to determine the tangent to the curve in the point of the last measurement d_0 (Fig.1). The angular coefficient of the tangent line k is:

$$k = 2ad_0 + b \quad (3)$$

The angular coefficient k of the tangent line is used for determination of the parameter of the measurement model (eq.1). The setting of the saturation limit K (eq.1) can be determined based on the statistical analysis of the wind data as the average value of the wind magnitude change in the chosen distance horizon.

Once the saturation limit K is selected, the distance D is then evaluated as:

$$D = \frac{K}{k} \quad (4)$$

Where D is the distance from the current aircraft position d_0 to the point where the tangent line defined by angular coefficient k intersects the line defined by saturation limit K .

deviation). The standard deviation of prediction at the point d_0 is assumed to be zero¹.

¹ The uncertainty associated to wind sensing is omitted for simplicity.

The linear function for the measurement standard deviation is selected as the Standard Deviation Model (Fig.2). At the distance of the horizon H , the value of the measurement model standard deviation $\sigma_{Measure}$ is equal to the value $\sigma_{Average}$. This

value represents the average standard deviation value of the average wind trend σ_{Trend} (the historical data). The value of the prediction horizon is determined on the basis of the RUC database statistics [5].

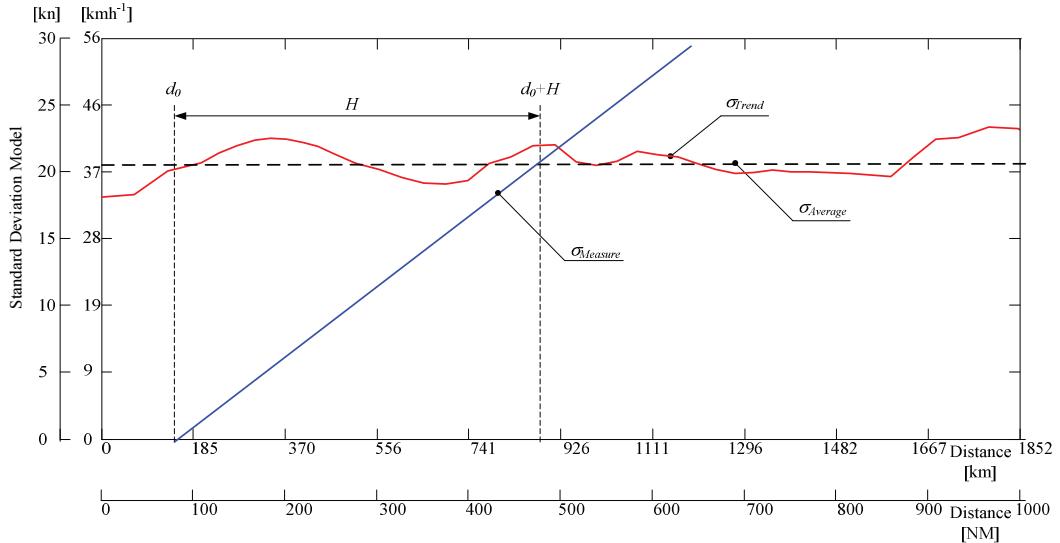


Fig.2. Standard Deviation Model

The measurement standard deviation model is:

$$\sigma_{Measure} = d \frac{\sigma_{Average}}{H} \quad (5)$$

where:

$\sigma_{Measure}$... the standard deviation associated to the extrapolation of the measurement,

$\sigma_{Average}$... the average standard deviation of the historical data.

C. Blending Algorithm

The prediction of the wind magnitude or bearing for the individual point of the flight plan is created by combination of the measurement prediction with the average trend value derived from weather database. Both of the values are weighted in accordance with their standard deviations [1]. The resulting value of the wind magnitude prediction is then:

$$\begin{aligned} Mag_{Prediction} &= \frac{\sigma_{Measure}^2}{\sigma_{Measure}^2 + \sigma_{Trend}^2} Mag_{Trend} + \\ &+ \frac{\sigma_{Trend}^2}{\sigma_{Measure}^2 + \sigma_{Trend}^2} Mag_{Measure} \end{aligned} \quad (6)$$

where the value $Mag_{Prediction}$ represents the resultant value of the wind magnitude measurement prediction. The value of the bearing is computed in the same manner.

III. ALGORITHM IMPLEMENTATION AND SETTING

The proposed algorithm was implemented and tested in the Matlab and FMS. During the implementation and testing phase some simplifications have been done. The simplification consists in the fixed setting of the saturation limits².

For the testing purposes and the evaluation the K and H coefficients of the measurement model have been preset to the given value (the coefficients have not been changed automatically). These values of the coefficients were:

- The saturation limit K corresponds to the 20% of the Mag_{d_0} : $K = 0.2 Mag_{d_0}$,
- The prediction horizon $H = 741$ km (400 NM).

A. Selection of the Scenarios

The selection of the test scenario is based on the analysis of the RUC weather grid database which includes the tracks of hourly generated weather reports and weather forecasts over US. Analysis of the weather grid defined suitable realistic wind scenarios to test the enhanced prediction algorithm.

The study [1] describes the scenario which represents average weather (i.e. the wind pattern with the highest probability of occurrence) and the extreme weather scenario (i.e. the wind pattern with the highest differences between wind values on grid).

² The value of the saturation limit is correlated with the wind dynamic behavior and it can be set automatically based on the analysis of results.

For the demonstrative purpose of the presented paper "One month scenario" is presented. The scenario was created for the comparison of the prediction algorithm behavior while the input conditions were selected from the period of several days during the year. Compact, one month period (28 days) year with the wind pattern about a noon was used for the testing.

B. Analysis Methodology

The performance of the presented algorithm (EWPA) was assessed by the comparison with the standard prediction algorithm (SPA) in absence of the wind forecast. The simulated flights have been conducted along the selected trajectory (i.e. CRUISE phase of flight between Los Angeles and Minneapolis).

The analysis of the results through 'one month scenario' was processed by the stochastic analysis of the complete set of the Trajectory Prediction errors from the single days (one month scenario). For the assessment of the prediction quality of the SPA and EWPA algorithms, the following metrics were used:

- *Mean* - Mean value of the absolute difference between the measured (real) and the predicted wind magnitude (bearing) values.
- *MSE* - Mean Squared Error of the difference between the measured (real) and the predicted wind magnitude (bearing) values.
- *Delay* - Delay on a selected time horizon caused by the worst orientation of the wind prediction error. Impact of the error in the wind magnitude is calculated for the pure head-wind direction. Impact of the error in the wind bearing is calculated from the bearing error of pure 100 knots strong cross-wind towards head-wind³.

C. Analysis results

The results of the algorithms comparison for wind magnitude are presented in the Table 1.

TABLE I. ALGORITHMS COMPARISON RESULTS

Prediction Time [min]	EWPA			SPA		
	MSE [km ² h ⁻²]	Mean [kmh ⁻¹]	Delay [s]	MSE [km ² h ⁻²]	Mean [kmh ⁻¹]	Delay [s]
8	4.1	1.5	0.9	32.6	4.3	2.4
15	27.8	3.9	4.3	106	7.6	8.3
20	68.6	6.1	9	182	9.8	14.4
30	223	11.3	24.7	353	13.5	29.9
60	962	22.2	99.5	806	20.4	91.1

Improvement on the shorter time horizons (i.e. 8min, 15min, and 20min) in the case of the Enhanced algorithm follows from figure 3. The reduced variability of the errors in the short time horizons indicates better adaptation to the natural dynamics of the wind (table I, MSE).

The box plot format of the figures allows an illustration of the evolution of the extreme and percentile values according to

³ The selected speed is M0.78.

the prediction time. The EWPA algorithm performance is better than the SPA performance in the short time horizon (8-20 minutes). The results for both algorithms are comparable also for the longer time horizon (30-60 minutes).

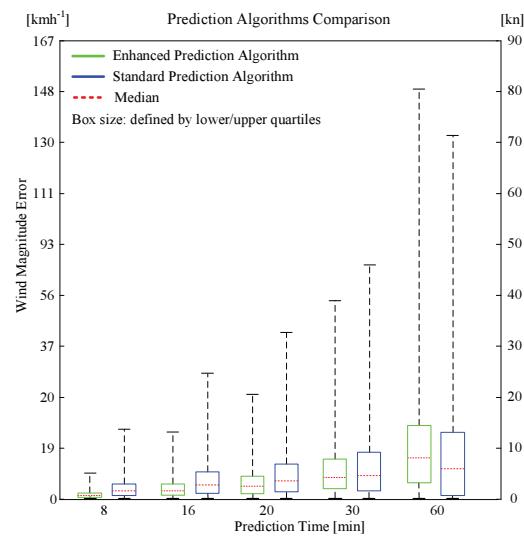


Fig.3 Algorithm Comparison

IV. CONCLUSION

The presented example indicates the EWPA ability to significantly improve trajectory prediction for the cruise level flight by the enhanced processing of the measured data used in a combination with the known historical wind trends along a selected trajectory. For longer prediction time horizons (above 1 hour) the algorithm EWPA cannot match the methods based on the presence of the actual up-to-date wind forecast, nevertheless numerous ATM applications using the short time are in use.

The proposed EWPA algorithm shows better prediction accuracy and precision, mainly for the short prediction horizons (up to 20 minutes), than the SPA.

The performance comparison of the EWPA algorithm with the methods using the actual up-to-date wind forecast is planned for the future work.

REFERENCES

- [1] Erasmus Project Report, D1.2.2 - Trajectory Prediction Enhancements, www.atm-erasmus.com, pp.76-84, December 2009
- [2] R. Siegwart, I. Nourbaksh, "Introduction to Autonomous Mobile Robots", A Bradford Book, MIT Press, 2004, ISBN 0-262-19502-X
- [3] E. Gelnarová, P. Krupanský, J. Svoboda, P. Mejzlík, T. Neužil, "Historical data based wind segmentation", MOSATT 2009, Kosice - Slovak Republic,
- [4] SESAR Consortium, SESAR Master Plan – D5, European Commission and Eurocontrol 2008, [cited 1 August, 2009]
- [5] B. E. Schwartz, S. G. Benjamin, S. M. Green, M. R. Jardin, "Accuracy of RUC-1 and RUC-2 Wind and Aircraft Trajectory Forecasts by Comparison with ACARS Observations", June 2000
- [6] C. R. Spitzer, R. Walter, "The Avionics Handbook", Chapter – "Flight Management Systems", CRC Press LLC., 2001, ISBN 0-8493-8348-X
- [7] K. Ogata, "Modern Control Engineering 3rd edition", Prentice Hall, 1997, ISBN 0-13-227307-1

Radar Cross Section Generation of the Possible Non-cooperative Targets

Rudolf Palme

Department of Aircraft and Ships

Budapest University of Technology and Economics

Budapest, Hungary

palme@rht.bme.hu

Abstract—This paper defines non-cooperative targets and presents a calculation method of their radar cross section also in bistatic cases. First a three-dimensional model is given which is built up by triangular facets. From the available flight path the azimuth and elevation angles are determined to position the target model. Then simple shadowing algorithm is used in order to have short computational time and acceptable accurate. The calculation of radar cross section uses physical optics theory. The surface integral is numerically evaluated over the illuminated surface only. Radar cross section is calculated by summarizing the results of each triangle. This work is part of the EU FP6 SINBAD (Safety and security Improved by New functionality for Better Awareness on airport approach and departure Domain) project leading by THALES.

Keywords: *Non-cooperative Target (NCT), TNB Frame, Triangular Facet, Shadowing, Physical Optics (PO), Radar Cross Section (RCS)*

I. INTRODUCTION

The number of Non-Cooperative Targets (NCT) is increasing with growth of aviation, widely use of Unmanned Aerial Vehicle (UAV), future operation of personal air vehicles and adaptation of birds to high density of air transport. It is clear that the non-cooperative targets may cause hazards and series emergency situations in air especially in airport regions. The needs in system can detect the NCTs and can predict possible hazard situations initiating by such targets are considerable increased especially after 9.11. The rate of the caused hazard depends on the target's technical parameters such as size, speed and maneuverability. Therefore it is highly recommended to classify NCTs. That was the reason why the European Commission started the Safety and security Improved by New functionality for Better Awareness on airport approach and departure Domain (SINBAD) project on 1st July, 2007. The scope for the SINBAD system covers the support of ATC Controllers (ATCOs) for the awareness of non-cooperative targets within the Controlled Terminal Region (CTR) and for the maintenance of the safety of air traffic (radar separation).

One of the main modules of SINBAD system is the NCT classification which aims to characterize the detected NCTs

against a set of aircraft classes. On the one hand, the classification method can be based on Radar Cross Section (RCS). Therefore the goal of this paper is developing RCS estimative algorithm at least to classify such targets but in optimal case to identify them.

II. NON-COOPERATIVE TARGETS

Those flying objects which do not communicate with the Air Traffic Control/Air Traffic Management (ATC/ATM) service are called as non-cooperative targets. This means that the aircraft is not responding to the interrogation signal transmitted by the Secondary Surveillance Radar (SSR), since it has no radar transponder, or if so it is switch off or having a system failure. So any flying object such as bird, UAV, hang glider, aircraft with failure in its communication system, small aircraft, terrorist's aircraft, etc. can possibly be a non-cooperative target.

In 2007, the number of the European non-cooperative targets is increasing [1]; the prediction to year 2020 is given in Fig. 1. The major reasons of this augmentation include the followings:

- the growth of the traditional air traffic that – having approximately equal system reliabilities – leads to a higher probability of an aircraft radio communication, or other fail that results in a non-cooperative flight,
- the enlargement of the unmanned aerial vehicles,
- the presents of the recreational flights (e.g. hang gliders, balloons),
- the appearance of the personal small aircraft (for a maximum of 6 passengers), which being designed to be accessible to common / ordinary people, might even be flown by pilots with limited experience,
- or the existence of birds at low altitudes.

The NCTs can cause hazards (decreasing in safety), conflicts and serious emergency situations (following with incidents or accidents). Therefore the new actual problem and task can be defined as the recognition, and identification of the non-cooperative targets and hazards in this way avoiding the emergency situations.

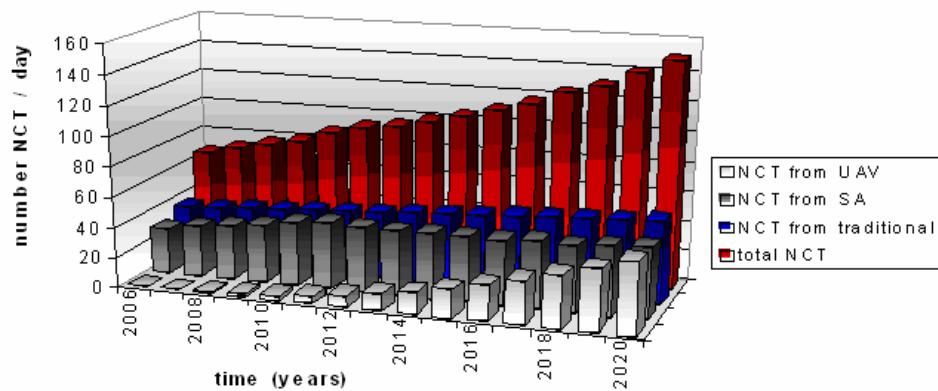


Figure 1. The evolution of the number of the European non-cooperative flights between 2006 and 2020 [1].

III. MODEL POSITIONING WITH USE OF FLIGHT PATH DATA

First of all some assumptions have been made, these are follows:

- the flight is a coordinated flight all the way,
- there is not any bank angles over 90 degrees in absolute value,
- the target behaves as a fixed wing aircraft,
- it flies with zero angle of attack.

The calculation based on the past of the flight which curve thus its derivative is given in Earth-axis system. Tangent, normal, and binormal unit vectors (\underline{T} , \underline{N} , \underline{B}) are calculated by using Frenet-Serret formulas. The vector \underline{tr} is pointing from the radar transmitter to the target. For model positioning it is necessary to determine the Euler angles of the target. In first step only the x-y coordinates are considered so a modified tangent unit vector (\underline{T}_{vy}) is also calculated. The angle between this (\underline{T}_{vy}) and the x direction vector is the yaw angle.

After that the TNB frame and \underline{tr} are rotated with negative yaw angle so \underline{T}_1 , \underline{N}_1 , \underline{B}_1 , and \underline{tr}_1 are given. Fig. 2 shows

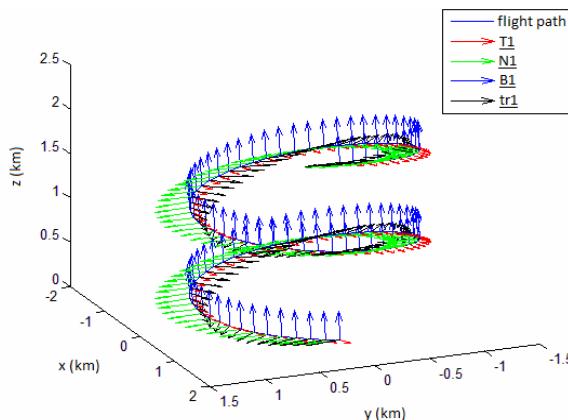


Figure 2. The flight path with rotated TNB frame.

these vectors on a helical flight path. In this situation the pitch angle is between \underline{T}_1 and the x direction vector.

Finally the bank angle is determined if the radius of curvature (R) and the target's speed (v) are known as

$$\text{bank} = \arctan\left(\frac{v^2}{R}\right). \quad (1)$$

Between the inputs of the RCS calculation there are the azimuth and elevation angles of the target from the radar transmitter/receiver viewpoint so it is also needed to calculate them. Now the \underline{T}_1 , \underline{N}_1 , \underline{B}_1 , and \underline{tr}_1 vectors are rotated with negative pitch angle, the results are \underline{T}_2 , \underline{N}_2 , \underline{B}_2 , and \underline{tr}_2 . Because the original TNB frame was not considered the bank position, another rotation is necessary which rotate \underline{tr}_2 with (1) so \underline{tr}_3 is given. The projection of the radar beam in the x-y plane is

$$\underline{\text{beam}}_x = \underline{tr}_3 - (\underline{tr}_3 * \underline{z}) * \underline{z} \quad (2)$$

where \underline{z} is the z direction vector. Afterwards the azimuth and elevation angles can be determined as

$$\phi = \pi - \arccos(\underline{T}_2 * \underline{\text{beam}}_x) \quad (3)$$

if \underline{tr}_3 is directed along positive direction of y axis, otherwise

$$\phi = \pi + \arccos(\underline{T}_2 * \underline{\text{beam}}_x), \quad (4)$$

$$\theta = \frac{\pi}{2} - \arccos(\underline{tr}_3 * \underline{\text{beam}}_x) \quad (5)$$

if \underline{tr}_3 is directed along negative direction of z axis, otherwise

$$\theta = \frac{\pi}{2} + \arccos(\underline{tr3} * \underline{beamx}). \quad (6)$$

IV. MODEL PREPARATION, SHADOWING

A. Model Preparation:

The models are given in Virtual Reality Modeling Language (VRML). VRML is a standard text file format for representing three-dimensional (3D) interactive vector graphics where, e.g., vertices and edges for a 3D polygon can be specified along with the surface color, UV mapped textures, shininess, transparency, and so on. Now those models are considered ones that are built up by triangular facets only. The text file describes the 3D geometrical shape but no material properties because the RCS calculation contains implicit the perfect conducting material property valid for the whole model surface. Fig. 3 shows a generic airliner in accordance with the principles above.

The current excited on the surface of the scatterer is found by the tangential components of the incident fields on the surface. Since the fields exist only on the illuminated portions of the scattering body, the PO current for a conducting body is given by

$$\underline{J}^s = 2\underline{n}_f \times \underline{H}^i \quad (7)$$

in the illuminated region where \underline{n}_f is the unit normal vector of the surface positioned outward and \underline{H}^i is the magnetic field vector. In the shadow region the PO current is zero [2]. Thus it is an important step to determine which triangle in which region stands.

B. Shadowing:

In order that the shadowing calculation be simple and fast some assumptions have been considered that are follows:

- Only the midpoint of the triangles is analyzed

instead of calculating overlap, because the surface of one triangle is negligible to the whole model.

- If there is a triangle between the analyzed midpoint and the radar transmitter, the triangle which owns the midpoint is located in the shadow region.

The calculation is made for every triangle. The first step is to define the position vector of the triangle's geometric center ($\underline{midpoint1}$). The vector of the radar beam (\underline{r}) is defined as the difference vector between $\underline{midpoint1}$ and the position vector of the radar transmitter. The nondirectional power density is – if the maximum of the transmitted power ($peakpower$) is given – as in [3]:

$$power = \frac{peakpower}{4\pi * \underline{r}^2}. \quad (8)$$

Now every triangle is considered in succession, until one is found which is located between the radar transmitter and the center of the analyzed triangle. If such a triangle exists, it is shaded, otherwise it is illuminated. The following analysis has to be performed for every triangle according to the next steps.

Label the vertices of the currently examined triangle to A, B and C. Constitute \underline{AB} , \underline{BC} and \underline{CA} vectors. Calculate the normal vector of the surface of the triangle, and sign it with \underline{n} . Define the \underline{u} and \underline{w} vectors as follows:

$$\underline{u} = -\underline{r}, \quad (9)$$

$$\underline{w} = \underline{midpoint1} - \underline{A} \quad (10)$$

where \underline{A} is the position vector of vertex A. Create the next scalars:

$$D = \underline{n} * \underline{u}, \quad (11)$$

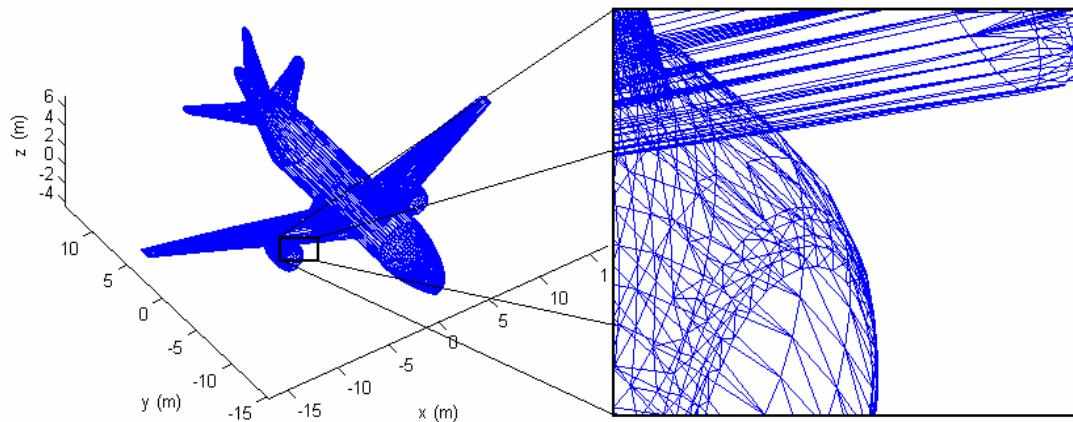


Figure 3. Generic airliner built-up by triangular facets.

$$N = -\underline{n}^* \underline{w}, \quad (12)$$

$$sI = \frac{N}{D}. \quad (13)$$

If the value (11) is close to zero (less than 10^{-7}), the surface of the analyzed triangle do not overlay the center of the other triangle. Otherwise the section of the analyzed dots intersects the surface if sI is between 0 and 1. In this case further analysis is needed to decide that the intersection is inside or outside the triangle because only inside case means shadowing. The position vector of intersection is

$$\underline{\text{Intpoint}} = \underline{\text{midpoint1}} + sI^* \underline{u}. \quad (14)$$

Mark the sections between the A, B and C vertices and the dot of intersection with \underline{AP} , \underline{BP} and \underline{CP} . So the intersection is inside the triangle if the next inequalities are true according to [4]:

$$0 < \underline{n}^* (\underline{AB} \times \underline{AP}), \quad (15)$$

$$0 < \underline{n}^* (\underline{BC} \times \underline{BP}), \quad (16)$$

$$0 < \underline{n}^* (\underline{CA} \times \underline{CP}). \quad (17)$$

If overlaying has not been found yet, eventually it is necessary to analyze that the intersection is on the edge of the triangle (or have a maximum distance of 10^{-4} m), and if it is there, it means shadowing also. First create $\underline{v1}$, the direction vector of the edge AB. Then define $\underline{pr1}$ as the vector between the dot of intersection (14) and a point of the examined edge (e.g. \underline{A}). The distance of the edge's line and the dot of intersection is

$$\underline{tav1} = |(\underline{A} + (\underline{pr1} * \underline{v1}) * \underline{v1}) - \underline{\text{Intpoint}}|. \quad (18)$$

If (18) less than 10^{-4} m it is necessary to examine that the section which contains the dot of intersection and which is perpendicular to the line of the edge intersects it between the two vertices or not. Shadowing is considered only the former case. The examination can be done as follows: In the equation of the edge's line the parameter of vertex A is known – it is zero – so only the parameter of vertex B is calculated as

$$t1 = \frac{\underline{B} - \underline{A}}{\underline{v1}}. \quad (19)$$

It means shadowing if the next inequalities are true:

$$0 \leq \underline{pr1}^* \underline{v1} \leq t1. \quad (20)$$

The analysis (18)-(20) is performed with the BC and CA edges also.

V. COORDINATE TRANSFORMATIONS [5]

Every triangle is located arbitrarily in a global coordinate system (x, y, z). In order to have the equations less complicated, a local coordinate system (x_l , y_l , z_l) can be defined to each triangle with vertices A, B, C in order according to [5]. Let the triangle lie on the $x_l - y_l$ plane in the newly defined local coordinate system. Put the origin of the local coordinate system at vertex A. Edge \underline{CA} has been taken along y_l axis. In this case the local coordinates are found using the following equations:

$$\underline{y}_l = -\frac{\underline{CA}}{|\underline{CA}|}, \quad (21)$$

$$\underline{z}_l = \frac{\underline{AB} \times (-\underline{CA})}{|\underline{AB} \times \underline{CA}|}, \quad (22)$$

$$\underline{x}_l = \underline{y}_l \times \underline{z}_l. \quad (23)$$

The vector \underline{c}_l represents the distance between the global and local coordinate centers. Thus it is equal to \underline{A} . Now the transformation matrix can be defined for the global coordinate system to the local rectangular coordinate system as

$$\underline{\underline{m}}_1 = \begin{bmatrix} \underline{x}^* \underline{x}_l & \underline{x}^* \underline{y}_l & \underline{x}^* \underline{z}_l \\ \underline{y}^* \underline{x}_l & \underline{y}^* \underline{y}_l & \underline{y}^* \underline{z}_l \\ \underline{z}^* \underline{x}_l & \underline{z}^* \underline{y}_l & \underline{z}^* \underline{z}_l \end{bmatrix}. \quad (24)$$

The parametric expressions for edges AB and BC can be written in local coordinates as the following:

$$\alpha(x_l) = \alpha_0 + \alpha_1 x_l \quad (25)$$

$$\beta(x_l) = \beta_0 + \beta_1 x_l \quad (26)$$

Fig. 4 explains the labels in the equations above. Since vertex A is the origin of the local coordinate system, the parameters in (25) are given as

$$\alpha_0 = 0, \quad (27)$$

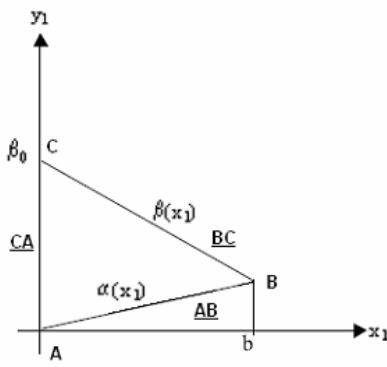


Figure 4. The triangle in local coordinates.

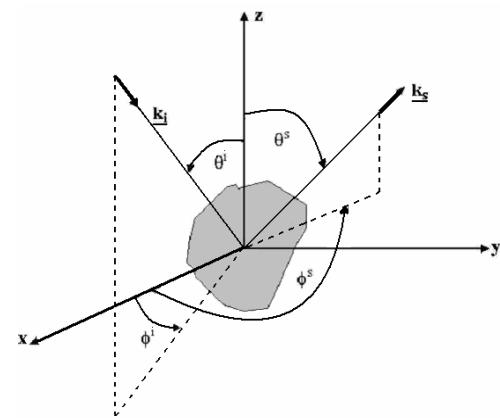


Figure 5. The scattering configuration.

$$\alpha_1 = \frac{AB_{ly}}{AB_{lx}} \quad (28)$$

if $AB_{lx} \neq 0$ otherwise

$$\alpha_1 = 0 \quad (29)$$

where $\underline{AB}_l = \underline{AB} * \underline{m}_1$, furthermore \underline{AB}_l indexing x and y refer to the first and second coordinate of the vector. Likewise the parameters in (26) are

$$\beta_0 = |\underline{CA}_l|, \quad (30)$$

$$\beta_1 = -\frac{\underline{BC}_{ly}}{|\underline{BC}_{lx}|}. \quad (31)$$

The incident field given in the global rectangular coordinates should also be transformed into the local coordinate system with the transformation matrix $\underline{\underline{m}}_1$. It is necessary to transform the incident field from the rectangular local coordinate system to the spherical local coordinate system, in which calculations for the scattered field will be done. Fig. 5 shows the scattering configuration in global coordinate system.

Because the origins of the local rectangular and spherical coordinate systems are the same, the transformation matrix and its building-up vectors are found to be

$$\underline{\theta}_l^i = \begin{bmatrix} \cos \theta_l^i \cos \phi_l^i \\ \cos \theta_l^i \sin \phi_l^i \\ \sin \theta_l^i \end{bmatrix}, \quad (32)$$

$$\underline{\phi}_l^i = \begin{bmatrix} -\sin \phi_l^i \\ \cos \phi_l^i \\ 0 \end{bmatrix}, \quad (33)$$

$$\underline{\underline{m}}_2 = \begin{bmatrix} \theta_l^i * \underline{x}_l & \phi_l^i * \underline{x}_l \\ \theta_l^i * \underline{y}_l & \phi_l^i * \underline{y}_l \\ \theta_l^i * \underline{z}_l & \phi_l^i * \underline{z}_l \end{bmatrix}. \quad (34)$$

where θ_l^i and ϕ_l^i are the incident azimuth and elevation angles in local coordinates. In case of the backscattered beam the calculation is the same as above except for the angles of incidence are replaced by the angles of scattering:

$$\underline{\underline{m}}_3 = \begin{bmatrix} \theta_l^s * \underline{x}_l & \phi_l^s * \underline{x}_l \\ \theta_l^s * \underline{y}_l & \phi_l^s * \underline{y}_l \\ \theta_l^s * \underline{z}_l & \phi_l^s * \underline{z}_l \end{bmatrix} \quad (35)$$

VI. PHYSICAL OPTICS (PO) FORMULATION

First the incident electric field is calculated over the illuminated triangles one after the other. The electric field density is given after [6] by

$$einc = \sqrt{\eta * power} \quad (36)$$

where power comes from (8) and η is the intrinsic impedance of free space.

$$\eta = 120\pi \Omega \quad (37)$$

The propagation vector is given in spherical coordinates by (see Fig. 5)

$$\underline{k}_i = -(\underline{x} \sin \theta^i \cos \phi^i + \underline{y} \sin \theta^i \sin \phi^i + \underline{z} \cos \theta^i). \quad (38)$$

If the source illuminating the target is at a far enough distance, then the incident field can be taken as a plane wave. The incident electric and magnetic fields are given by the following expressions by [5]:

$$\underline{E}^i = \underline{E}_0^i e^{-j*k*\underline{k}_i*\underline{r}} \quad (39)$$

where \underline{k}_i comes from (38), \underline{E}_0^i is the value of (36) with the direction \underline{k}_i and k is the wave number:

$$k = \frac{2\pi f}{c} \quad (40)$$

where c is the speed of light in vacuum and f is the frequency of the wave.

$$\underline{H}^i = \underline{H}_0^i e^{-j*k*\underline{k}_i*\underline{r}} \quad (41)$$

where

$$\underline{H}_0^i = \frac{1}{\eta} \underline{k}_i \times \underline{E}_0^i \quad (42)$$

To the RCS calculation some assumptions have been made these are summarized as follows:

- the solid models are built up from finite triangles,
- the whole model is created from the same conducting material,
- the models have rigid body,
- the distance between the radar and the target is large so the radar signals are perpendicular to each other and
- the analysis is static.

Thus the PO current for the model surface is given by (7). A plane wave of arbitrary polarization is incident from an angle (θ_l^i, ϕ_l^i) . The wave polarization is determined by the constants $E_{\theta l}^i$ and $E_{\phi l}^i$ in the expression

$$\underline{E}_l^i = (E_{\theta l}^i \underline{\theta}_l^i + E_{\phi l}^i \underline{\phi}_l^i) e^{-j*k*\underline{k}_l^i*\underline{r}_l} \quad (43)$$

where \underline{r}_l is the position vector of the even examined facet in spherical coordinates. The surface current induced on the $+z$ side of the facet is given as

$$\begin{aligned} \underline{J}_l^s = & \frac{2e^{-jk\underline{k}_l^i \cdot \underline{r}_l}}{\eta} (\underline{x}_l (\cos \phi_l^i E_{\theta l}^i - \cos \theta_l^i \sin \phi_l^i E_{\phi l}^i) + \\ & + \underline{y}_l (\sin \phi_l^i E_{\theta l}^i + \cos \theta_l^i \cos \phi_l^i E_{\phi l}^i)) \end{aligned} \quad (44)$$

According to [7] if the position vector to a source point is \underline{p}' and the unit vector in the direction of the observation point is

$$\underline{p} = u\underline{x} + v\underline{y} + w\underline{z} \quad (45)$$

where u, v, w are the direction cosines of the observation point:

$$u = \sin \theta \cos \phi, \quad (46)$$

$$v = \sin \theta \sin \phi, \quad (47)$$

$$w = \cos \theta; \quad (48)$$

then the components of the electric field tangential to a sphere at radius r , i.e. the radiation integrals are

$$\begin{aligned} E_{\theta}^s(x, y, z) = & \frac{-jk\eta}{4r\pi} e^{-jkr} * \\ & * \iiint_V (J * \underline{\theta} + \frac{1}{\eta} * J_m * \underline{\phi}) e^{jk\underline{p}' \cdot \underline{p}} dV' \end{aligned} \quad (49)$$

$$\begin{aligned} E_{\phi}^s(x, y, z) = & \frac{-jk\eta}{4r\pi} e^{-jkr} * \\ & * \iiint_V (J * \underline{\phi} - \frac{1}{\eta} * J_m * \underline{\theta}) e^{jk\underline{p}' \cdot \underline{p}} dV' \end{aligned} \quad (50)$$

where J_m is the magnetic current and it is zero considered the assumptions. Using (44) in the (49) radiation integral gives with $J_m = 0$ in the source point (x_l, y_l)

$$\begin{aligned} E_{\theta}^s(x_l, y_l) &= \frac{-jk\eta}{4\pi|r|} e^{-jk|r|} \iint_s \frac{2e^{jkh}}{\eta} * \\ &* [(E_{0\theta} \cos \phi_l^i - E_{0\phi} \cos \theta_l^i \sin \phi_l^i) \cos \theta_l^s \cos \phi_l^s + \\ &+ (E_{0\theta} \sin \phi_l^i + E_{0\phi} \cos \theta_l^i \cos \phi_l^i) \cos \theta_l^s \sin \phi_l^s] e^{jkg} dx_l dy_l \end{aligned} \quad (51)$$

where

$$h = x_l \sin \theta_l^i \cos \phi_l^i + y_l \sin \theta_l^i \sin \phi_l^i, \quad (52)$$

$$g = x_l \sin \theta_l^s \cos \phi_l^s + y_l \sin \theta_l^s \sin \phi_l^s. \quad (53)$$

With the assumptions (51) is

$$\begin{aligned} E_{\theta}^s(x_l, y_l) &= \frac{-jk\eta}{4\pi|r|} e^{-jk|r|} \frac{2}{\eta} \iint_s e^{jkh} * e^{jkg} dx_l dy_l * \\ &* [(E_{0\theta} \cos \phi_l^i - E_{0\phi} \cos \theta_l^i \sin \phi_l^i) \cos \theta_l^s \cos \phi_l^s + . \quad (54) \\ &+ (E_{0\theta} \sin \phi_l^i + E_{0\phi} \cos \theta_l^i \cos \phi_l^i) \cos \theta_l^s \sin \phi_l^s] \end{aligned}$$

Similarly

$$\begin{aligned} E_{\phi}^s(x_l, y_l) &= \frac{-jk\eta}{4\pi|r|} e^{-jk|r|} \frac{2}{\eta} \iint_s e^{jkh} * e^{jkg} dx_l dy_l * \\ &* [(-E_{0\theta} \cos \phi_l^i + E_{0\phi} \cos \theta_l^i \sin \phi_l^i) \sin \phi_l^s + . \quad (55) \\ &+ (E_{0\theta} \sin \phi_l^i + E_{0\phi} \cos \theta_l^i \cos \phi_l^i) \cos \phi_l^s] \end{aligned}$$

These surface integrals can be analytically calculated over the triangular facets. The integral in itself is given as

$$I_0 = \int_{x_l=0}^b \int_{y_l=\alpha(x_l)}^{\beta(x_l)} e^{jk((\sin \theta_l^i \cos \phi_l^i + \sin \theta_l^s \cos \phi_l^s)x_l + \\ + (\sin \theta_l^i \sin \phi_l^i + \sin \theta_l^s \sin \phi_l^s)y_l)} dx_l dy_l. \quad (56)$$

The limits of the integration are the edges of the particular triangle. The expressions for the integral limits are (25)-(26) with constants (27) – (31) and

$$b = AB_{lx}. \quad (57)$$

The scattered field from a single triangle is given as

$$\underline{E}_l^s = \underline{E}_{\theta l}^s \underline{\theta}_l + \underline{E}_{\phi l}^s \underline{\phi}_l. \quad (58)$$

The RCS formula from [2] is used to compute the RCS of the target:

$$\sigma = \lim_{r \rightarrow \infty} 4\pi r^2 * \frac{|\underline{E}_l^s|^2}{|\underline{E}_l^i|^2}. \quad (59)$$

The RCS results from each illuminated triangle are summed up in order to find the total radar cross section. Fig. 6 shows the ideal and calculated monostatic scattering pattern for square with vertical polarization and with edge lengths of 5λ where λ is the wavelength. Usually the frequency-independent quantity $\frac{\sigma}{\lambda^2}$ is plotted rather than σ .

Because the computational method is under development, it produces analytical errors in case of some special configuration (see on Fig. 6 the calculated graph) which are

- $\sin \theta_l^i \sin \phi_l^i + \sin \theta_l^s \sin \phi_l^s =$
 $= \sin \theta_l^i \cos \phi_l^i + \sin \theta_l^s \cos \phi_l^s$ and $\beta_1 = -1$
- or $\beta_1 = \alpha_1$.

It is possible to avoid these errors with another better coordinate transformation but the further examination is not scope of this paper. Besides the calculation method needs to accelerate, there is not any calculation with complex models which takes up a lot of time, only results from a hypothetical aircraft are given in Fig. 7.

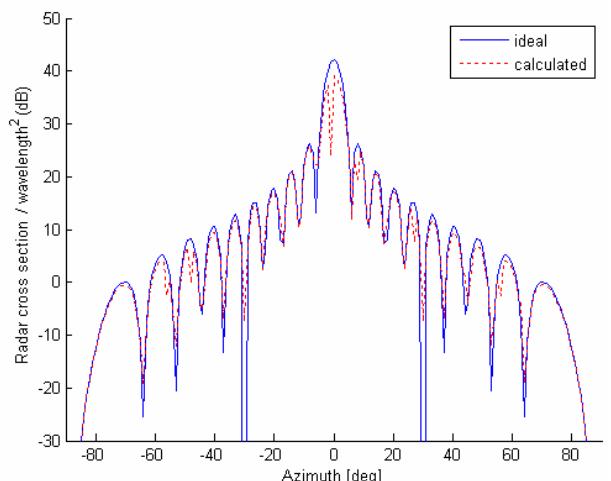


Figure 6. Ideal [7] and calculated radar cross section of square plate ($\phi = 0$ deg).

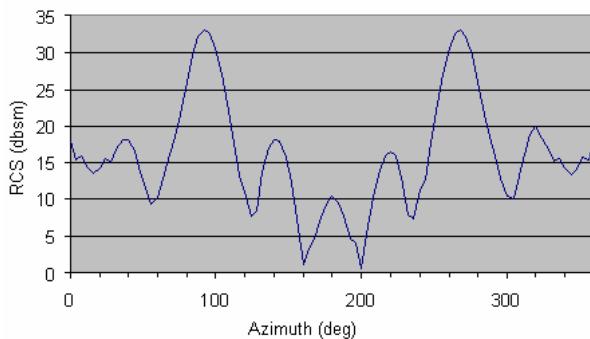


Figure 7. RCS diagram of a hypothetical aircraft at 10 GHz ($\phi = 0$ deg).

VII. IN CONNECTION WITH SINBAD PROJECT

The three new services provided by SINBAD Active Hazard Assessment (AHA) module may be summarized as follows [8]:

- Automated support for the detection of airspace infringements by NCT through appropriate notification of concerned ATCO;
- Improvement to ATCO situational awareness about NCT (with altitude and classification);
- Automated support for the notification of Security Units about NCT creating a potential threat to ATM security.

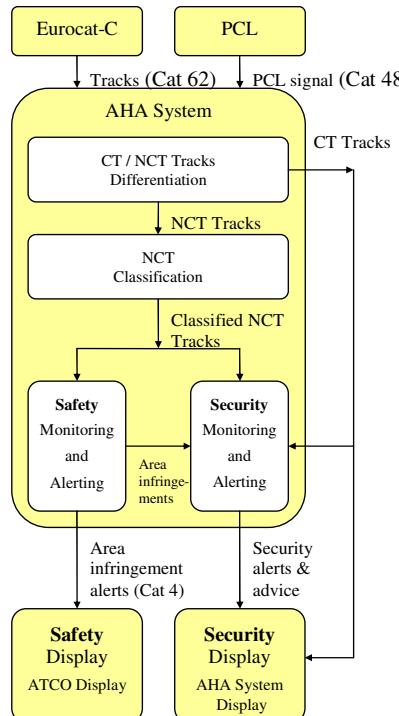


Figure 8. AHA global functional architecture

The main objective of the AHA system is to assist the air traffic controller and the ATM security officer respectively by providing safety alerts and security alerts. The safety alerts comprise of area infringement alerts when NCT's are entering the CTR. The security alerts comprise of area infringement alerts when NCT's are entering the CTR supplemented by information about potential NCT intentions, threat levels and advice about potential actions to assist the ATM security officer. An overview of the AHA functional architecture is presented in Figure 8.

VIII. CONCLUSION

The objective of this paper was to provide radar cross section calculation software to support the investigations of the EU sponsored SINBAD project, and to establish the background for further relevant research. Three dimensional VRML models were created built-up by triangular facets. The calculation based on these models as perfectly conducting solid models. Then the flight path was analyzed the Euler, azimuth and elevation angles were determined. To each triangle a ray tracing algorithm were applied that was developed to reach simple and fast computation time with acceptable accurate. PO integrals were formulated for each illuminated triangle which were calculated analytically. The radar cross section was summed up to get bistatic RCS from all possible viewpoint. The method presented in this paper is used by SINBAD Active Hazard Assessment System to classify NCTs. The accurate of the results will be tested in Brno TMA (Terminal Control Area) with radar measurements in the summer of 2010 as scheduled.

The calculated data were reliable; the applied assumptions have imperceptible influence to the results. Note that the method needs further relevant development but the NCT classification is achievable with suitable 3D models and with calculation several different bistatic RCS pairs.

REFERENCES

- [1] Dániel Rohács, "Prediction of the non-cooperative air targets at the European airport vicinities," Internal SINBAD Project Progress Report, unpublished.
- [2] Eugene F. Knott, John F. Shaeffer, Michael T. Tuley, "Radar cross section," Chapter 3, Artech House, 1985.
- [3] Merv C Budge, "Intro to radar systems," Chapter: Radar Range Equations, The University of Alabama in Huntsville, <http://www.ece.uah.edu/courses/material/EE619/index.htm>, 2005.
- [4] Iványi Antal (editor), "Algorithms of informatics," Chapter: Computer Graphics, Eötvös Kiadó, 2004.
- [5] N. A. Albayrak, "RCS computations with PO/PTD for conducting and impedance objects modelled as large flat plates," Bilkent University, unpublished.
- [6] John D. Jackson, "Classical electrodynamics third edition," John Wiley & Sons Inc., New York, 1998.
- [7] David C. Jenn, "Radar and laser cross section engineering second edition," Chapter 2, American Institute of Aeronautics and Astronautics Inc., Reston, 2005.
- [8] Emmanuel Isambert, Olivier Desenfans, Jean-Pierre Lesueur, "Operational Concept Document," Internal SINBAD Project OCD, unpublished.

Study on Conflict Detection Method with Downlink Aircraft Parameters

Atsushi SENOGUCHI, and Yutaka FUKUDA

Air Traffic Management Department
Electronic Navigation Research Institute
Tokyo, Japan
senoguchi@enri.go.jp, fukuda@enri.go.jp

Abstract—In order to maintain the safety of air traffic, ENRI (Electronic Navigation Research Institute) tries to improve CA (Conflict Alert) functionality using airborne information via Mode S data-link. CA supports air traffic controller in their maintaining of safe separation between aircraft by predicting aircraft positions and detecting potential conflicts. The aim of this study is to develop a new DAPs-CDM (Conflict Detection Method using Downlink Aircraft Parameters) and evaluate the impact of its introduction by computer simulation.

Firstly, in order to understand the characteristics of the conventional CDM, we calculated horizontal and vertical prediction errors in aircraft position. The conventional CDM uses linear prediction with only radar information on the ground. We also analyzed CA occurrences on conventional CDM. We found that both horizontal and vertical prediction errors were reduced by using airborne information in addition to radar information. We also found that it was better to smooth vertical speed for prediction and to utilize selected altitude in DAPs-CDM.

Finally, the characteristics of DAPs-CDM were studied and the advantages were demonstrated. The new function of DAPs-CDM is to predict aircraft positions using aircraft velocity on the airborne side and to determine aircraft flight phases using roll angle and selected altitude. For the purpose of comparing DAPs-CDM with the conventional CDM, ENRI developed CDES (Conflict Detection Evaluation System). It can simulate both DAPs-CDM and the conventional CDM under air traffic situations and system parameters almost the same as operational situations and system parameters. As a result of computer simulation with CDES, the determination of vertical flight phases by selected altitude was most effective in reducing the number of unnecessary CAs.

Keywords-component; Conflict Alert; Conflict Detection Method; Downlink Aircraft Parameters

I. INTRODUCTION

From the perspective of safety, it is necessary to ensure spacing between aircraft. In Japan, the horizontal separation minimum is 5 NM (1 NM = 1,852 m) under radar control. The vertical separation minimum is 1,000 ft (1 ft = 0.3048 m) where altitude is below 41,000 ft under IFR (Instrument Flight Rules) and RVSM (Reduced Vertical Separation Minimum), or 2,000 ft where altitude is above 41,000 ft. [1]

A situation where the proximity distance between aircraft cannot be satisfied in terms of both the horizontal and vertical minima is called a conflict. In order to ensure proper spacing between aircraft, the air traffic controller issues instructions to aircraft as needed, based on the air traffic situation.

The RDP (Radar Data Processing System), which air traffic controllers use for en-route airspace in Japan, has the function of predicting aircraft positions and detecting potential conflicts, called CA (Conflict Alert). This function predicts aircraft positions 3 minutes into the future, using current aircraft positions and estimated aircraft velocities by a tracking process. When a potential conflict is detected, a warning message for the air traffic controller is displayed on the RDP screen. [2]

With this function, an aircraft is assumed to fly at constant speed. Therefore, when predicting aircraft positions, prediction errors can occur due to estimation differences and fluctuations in aircraft velocity. Being linear prediction, it is also difficult to predict changes in aircraft flight phases, such as from straight to turning or from climbing to cruising.

Currently, SSR (Secondary Surveillance Radar) Mode S, which is a new radar system for air traffic control, is being introduced. Mode S is able not only to get aircraft positions more accurately, but also to have the function of to downlink aircraft parameters such as ground speed, vertical speed, magnetic heading and selected altitude via air-to-ground digital communication. Having these airborne parameters available on the ground side is expected to contribute to advanced air traffic management. In terms of conflict detection, the usage of DAPs (Downlink Aircraft Parameters) will reduce the number of unnecessary CAs and air traffic controllers' workloads. [3]

The aim of this study is to develop a new DAPs-CDM (Conflict Detection Method using Downlink Aircraft Parameters) and evaluate the effectiveness of its introduction by computer simulation. [4]

Firstly, overviews of CA and CDM (Conflict Detection Method) are introduced in Section 2 and 3. Then, in order to better understand CDM, the results of analyzing prediction error on DAPs-CDM are explained in Section 4. Moreover, for the same purpose, analyses of situations in which CA occurred are explained in Section 5. Finally, in Section 6, the characteristics of DAPs-CDM are studied and the results of

simulation which verify the effectiveness of its introduction are demonstrated.

II. TRENDS IN CA AND MODE S

Trends in STCA (Short Term Conflict Alert) and SSR Mode S are introduced.

A. STCA Trends

Reference [5], published by ICAO (International Civil Aviation Organization) in 2001, describes the procedures of STCA; e.g., “*A statistical analysis should be made of justified alerts in order to identify possible shortcomings in airspace design and ATC procedures as well as to monitor overall safety levels.*”

EUROCONTROL developed the operational requirements of STCA in 1998. [6] For the purpose of improving air traffic safety, the European SSAP (Strategic Safety Action Plan) was put in place in 2003. The SSAP was followed up by the ESP (European Safety Programme) for ATM in 2006. The ESP was for the improvement of safety in 5 main fields, one of them being system safety nets, which the ESP aimed to consolidate. STCA was considered to be a safety function in on-the-ground systems.

Because the functions and procedures of STCA were not standardized across European countries, the specifications of STCA were set, and guidelines devised. [7][8] Policy, training, regulation and requirement are described in the specifications from the viewpoint of management. The guidelines are more concerned with the technical aspects of STCA.

Efforts toward safety improvement are being conducted by STCA with its SESAR Master Plan using enriched surveillance information. [9]

B. Mode S Trends

The international standard of BDS (Comm-B Data Selector) code related with the data format of SSR Mode S was implemented in [10], published by ICAO in 2002.

In Europe, SSR Mode S ELS (Elementary Surveillance) was implemented, and Mode S EHS (Enhanced Surveillance) was implemented in the France, Germany, and England. [11]

- Required Capability of ELS

BDS 1,0: Data Link Capability

BDS 1,7: Common Usage GICB Capability

BDS 2,0: Aircraft Identification

- Required Capability of EHS

BDS 4,0: Selected Vertical Intention

BDS 5,0: Track and Turn

BDS 6,0: Heading and Speed

In Japan, the number of aircraft equipped with a transponder that is Mode S data-link capable is increasing. [12]

III. CONFLICT ALERT IN JAPAN

This section explains a functionality overview of CA and CDM in Japan. In 1979, CA was implemented as an RDP function. In 2002, CDM with flight plan information was added to the CA function. [13][14]

Fig. 1 shows an example of a CA display. Triangular symbols stand for aircraft positions and each symbol has an information tag. The information tags display a call sign in the first row, altitude information in the second row and ground speed in the third row. If the CA function detects potential conflicts up to 3 minutes in advance, it displays a blinking ‘CNF’ in the information tags of related aircraft and, at the same time, renews the CA status list. In Fig. 1, the list is shown in the upper left corner.

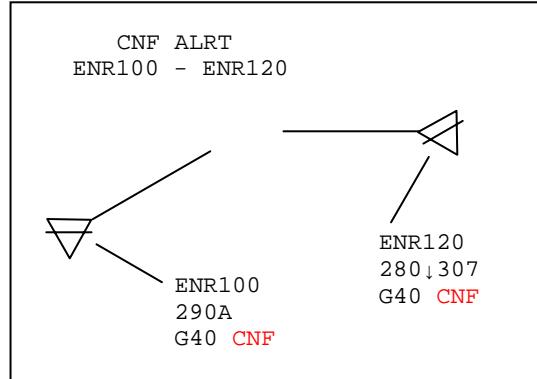


Figure 1. Example of CA Display

This section also describes how the CA function detects potential conflicts. The conventional CDM is divided into 2 types: the LP-CDM (Linear Prediction - Conflict Detection Method) and the FP-CDM (Flight Plan - Conflict Detection Method).

A. LP-CDM

Fig. 2 shows the concept of LP-CDM. As LP-CDM assumes that aircraft continue to fly at constant speed, predicted position $\mathbf{p}_i(t, \tau)$ is given by

$$\mathbf{p}_i(t, \tau) = \mathbf{x}_i(t) + \tau \cdot \mathbf{v}_i(t). \quad (1)$$

$\mathbf{x}_i(t)$ and $\mathbf{v}_i(t)$: Position and velocity of aircraft i

LP-CDM executes the checks of (2) and (3) for all the aircraft combinations (i, j) in flight time t , where time τ changes in $0 \leq \tau \leq T_p$. If both (2) and (3) are satisfied, LP-CDM regards the situation as a conflict.

$$\sqrt{(p_{xi}(t, \tau) - p_{xj}(t, \tau))^2 + (p_{yi}(t, \tau) - p_{yj}(t, \tau))^2} \leq R_h \quad (2)$$

$$|p_{zi}(t, \tau) - p_{zj}(t, \tau)| \leq R_v \quad (3)$$

$$\mathbf{p}_i(t, \tau) = (p_{xi}(t, \tau), p_{yi}(t, \tau), p_{zi}(t, \tau)) \quad (4)$$

In terms of LP-CDM on RDP, the observed position of aircraft is used as $\mathbf{x}_i(t)$ and estimated velocity of aircraft by tracking observed positions with alpha-beta filter is used as $\mathbf{v}_i(t)$. [15] Prediction time T_p is 3 minutes, horizontal detection threshold R_h is 5 NM, and vertical detection threshold R_v is 700 ft if altitude is below 41,000 ft and 1,600 ft if altitude is above 41,000 ft.

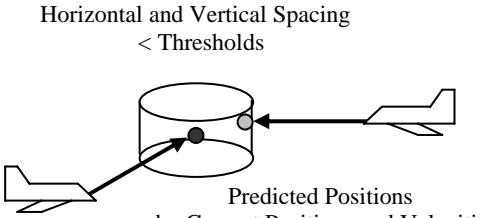


Figure 2. LP-CDM Concept

B. FP-CDM

Fig. 3 shows a concept of FP-CDM. In comparison with LP-CDM, FP-CDM uses flight plan information. The flight plan is submitted to the ATS (Air Traffic Service) provider before departure. The route the aircraft is going to fly is included in the flight plan information. FP-CDM checks whether or not aircraft are flying on the planned route. If an aircraft is flying on the route, FP-CDM extends the prediction course along the route. When the route bends, the prediction course bends at waypoints. On the other hand, if the aircraft isn't flying on the route, FP-CDM is equal to LP-CDM.

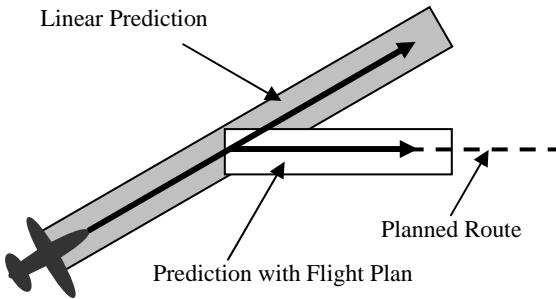


Figure 3. FP-CDM Concept

IV. ANALYSIS ON PREDICTION ERROR OF POSITION

In order to better understand CDM, prediction errors in aircraft position on (1) using flight recorded data were compared with those using radar data. In this section, we discuss horizontal and vertical prediction errors separately.

A. Horizontal Prediction Error

Observed positions and estimated velocities of aircraft by tracking observed positions on the ground side are recorded in radar data. The update rate for them in en route surveillance in Japan is every 10 seconds.

Aircraft positions and velocities (ground speed and true track angle) measured by airborne sensors are recorded in flight recorded data. The update rate for them is every 1 second in the case of new aircraft types.

Fig. 4 shows the tracks of aircraft flying from Haneda airport to Fukuoka airport in Japan. The map is plane-projected centered on the Tokyo Area Control Center. The unit of X and Y axis is the NM. In Fig. 4, aircraft fly to turn around airports. Far from airports, aircraft fly mostly straight. The horizontal aircraft profile was analyzed.

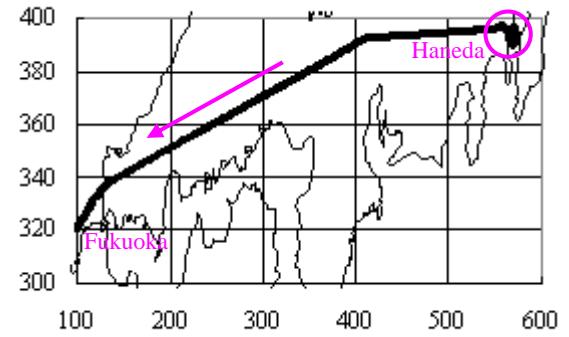


Figure 4. Aircraft Tracks

Horizontal prediction errors in position are defined as

$$E_h = \sqrt{(X_{xi}(t+T_p) - p_{xi}(t, T_p))^2 + (X_{yi}(t+T_p) - p_{yi}(t, T_p))^2}. \quad (5)$$

$X_{xi}(t+T_p)$ and $X_{yi}(t+T_p)$ are horizontal components of aircraft position $\mathbf{X}_i(t+T_p)$ in flight recorded data. $p_{xi}(t, T_p)$ and $p_{yi}(t, T_p)$ are horizontal components of predicted position $\mathbf{p}_i(t, \tau)$ where $\tau = T_p = 3$ minutes.

Ground speed indicates the absolute value of horizontal aircraft velocity and its unit is the knot (1 knot = 1,852/3,600 m/s). Aircraft flight phases based on ground speed are classified into acceleration, constant speed and deceleration phases. True track angle indicates the argument value of horizontal aircraft velocity and its unit is the degree. True North is 0 degree and positive rotation is clockwise. Aircraft flight phases based on true track angle are classified into straight and turning phases. Horizontal components of velocity ($v_{xi}(t), v_{yi}(t)$) are combined with ground speed $GS_i(t)$ and true track angle $TTA_i(t)$ by

$$\begin{pmatrix} v_{xi}(t) \\ v_{yi}(t) \end{pmatrix} = \begin{pmatrix} GS_i(t) \cdot \cos\left(\frac{\pi}{2} - \frac{\pi}{180} \cdot TTA_i(t)\right) \\ GS_i(t) \cdot \sin\left(\frac{\pi}{2} - \frac{\pi}{180} \cdot TTA_i(t)\right) \end{pmatrix}. \quad (6)$$

Fig. 5 shows horizontal prediction errors in position using radar data and flight recorded data. Calculation timing is every 10 seconds in accordance with the update rate for radar data.

Horizontal prediction error in position using flight recorded data is reduced in comparison with that using radar data. In straight phase, it is reduced by 37% on average (from 1.43 NM to 0.92 NM). In turning phase, it is reduced by 28% on average (from 6.51 NM to 4.67 NM).

When predicting horizontal aircraft positions, it is important to use airborne velocity (ground speed and true track angle) and to determine adequately whether the aircraft flight phase is in straight or turning.

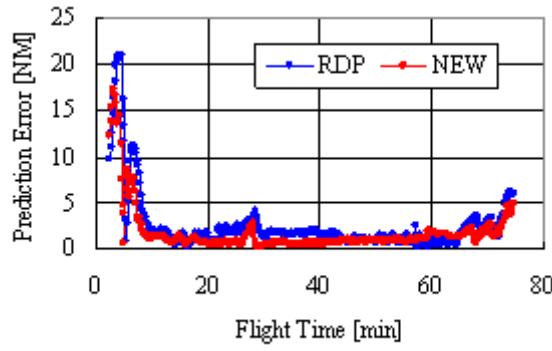


Figure 5. Horizontal Prediction Error

B. Vertical Prediction Error

Observed altitude of aircraft and estimated vertical rates of aircraft by tracking observed altitude on the ground side are recorded in radar data. The update rate for them in en route surveillance in Japan is every 10 seconds.

Aircraft altitude and vertical rates measured by airborne sensors are recorded in flight recorded data. The update rate for them is every 1 second in the case of new aircraft types.

Fig. 6 shows the altitude from climb through cruise to descend. The vertical aircraft profile was analyzed.

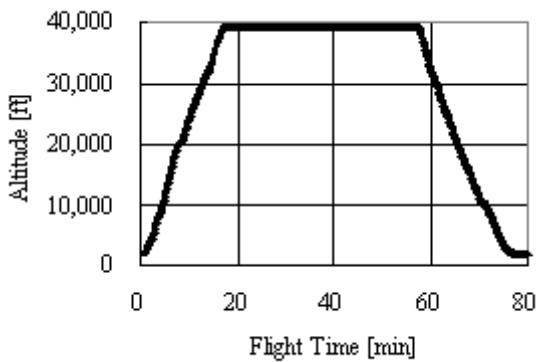


Figure 6. Aircraft Altitude

Fig. 7 shows the vertical rate of altitude. The vertical rate in cruise phase is a constant 0 ft/min, but fluctuates in climb and descend phases. In Fig. 7, the optimized vertical rate by alpha-beta filter in climb and descend phases is shown. Filter gain α is 0.05 ($\beta = 0.001282$), initial values are 3,000 ft/min in climb

phase and -3,000 ft/min in descend phase. Optimized vertical rate in climb and descend phase is smoothed for prediction.

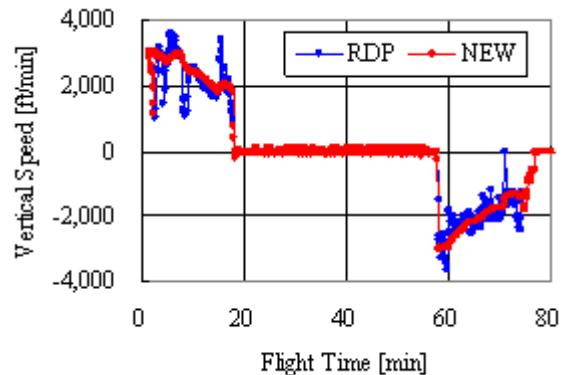


Figure 7. Altitude Vertical Rate

Vertical prediction error in position is defined as

$$E_v(t) = |X_{zi}(t+T_p) - p_{zi}(t, T_p)|. \quad (7)$$

$X_{zi}(t+T_p)$ is the vertical component of aircraft position $\mathbf{X}_i(t+T_p)$ in flight recorded data. $p_{zi}(t, T_p)$ is vertical components of the predicted position $\mathbf{p}_i(t, \tau)$ where $\tau = T_p = 3$ minutes.

Fig. 8 shows vertical prediction errors in position using a vertical rate of radar data and an optimized vertical rate. Calculation timing is every 10 seconds in accordance with the update rate for radar data. Vertical prediction errors in position using an optimized vertical rate in climb and descend phases are reduced in comparison with those using the raw vertical rate. Standard deviation of vertical prediction error is reduced by 23% (from 1,500 ft to 1,150 ft).

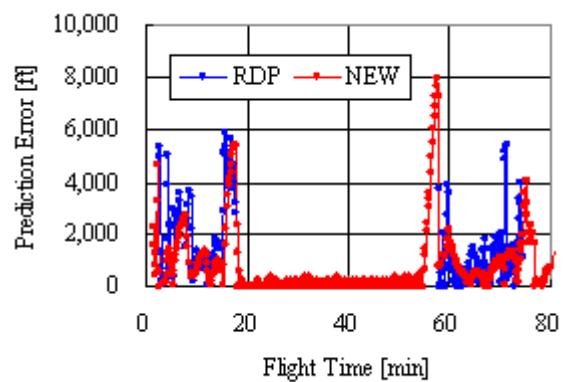


Figure 8. Vertical Prediction Error

In Fig. 8, there are large vertical prediction errors with changes in aircraft flight phase, such as from climb to cruise in the case of using both raw and optimized vertical rate. This means that it is impossible to predict the change in aircraft

flight phases by LP-CDM only. To reduce this vertical prediction error, it is necessary to use selected altitude built into LP-CDM.

V. ANALYSIS OF CA OCCURRENCE

In this section, we discuss how CA occurrences were analyzed to better understand CDM. [16]

A. Simulation Conditions

We simulated CA occurrences in the current operational environment. The air traffic scenario was constructed from radar data and flight plan data of the Tokyo Area Control Center. Air traffic volume consisted of 1 hour of peak time air traffic. The total number aircraft was 409.

The conditions of conflict detection were also decided on with reference to operational system parameters. The horizontal threshold R_h was set as 5 NM. The vertical threshold R_v was basically set as 700 ft and as 1,600 ft in the case of that altitude being over 41,000 ft.

B. Results

TABLE I shows the number of CA occurrences by simulation. The numbers of CA occurrences by LP-CDM and FP-CDM were 135 and 109 respectively. These numbers included the plural times caused by the same aircraft pair. 84 and 77 pairs of aircraft made all CA occurrences by LP-CDM and FP-CDM. 30 and 21 pairs of all made more than 2 times of CA occurrence.

TABLE I. NUMBER OF CA OCCURRENCES

		All	Only 1 Time	More than 2 Times
LP-CDM	Occurrences	135	54	81 (60%)
	Pairs	84	54	30 (36%)
FP-CDM	Occurrences	109	56	53 (49%)
	Pairs	77	56	21 (27%)

Fig. 9 shows that the frequencies of CA durations by LP-CDM and FP-CDM are the same trends, as both 75% of CA occurrences by LP-CDM and FP-CDM persisted for 0-20 seconds. Fig. 10 shows that the frequencies of CA intervals by LP-CDM and FP-CDM also are the same trends as 71% and 59% of CA occurrences by LP-CDM and FP-CDM respectively, repeated every 0-20 seconds, in terms of more than 2 CA occurrences. Considering the results of Fig. 9 and Fig. 10, intermittent CA might often occur with conventional CDM.

Fig. 11 shows the classification of vertical situations at the time of CA occurrence. The vertical axis represents the number of aircraft pairs. ‘0-’ stands for altitude differences between 0 ft and 5,000 ft. ‘50-’ stands for altitude differences between 5,000 ft and 10,000 ft. ‘100-’ stands for altitude differences above 10,000 ft. 3 left-hand bars are made from LP-CDM and 3 right-hand others are made from FP-CDM. Depending on vertical flight phases of aircraft at the time of CA occurrence, vertical situations are classified into 6 patterns; the combinations of climb ‘↑’, descend ‘↓’ and cruise ‘→’.

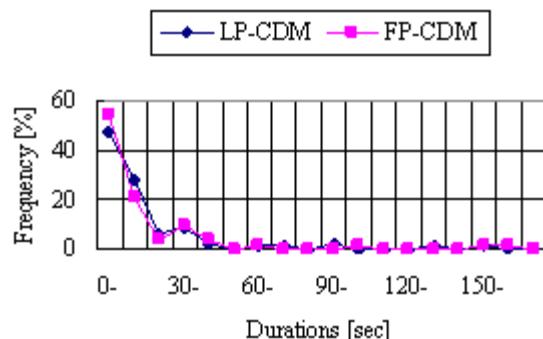


Figure 9. Frequency of CA Durations

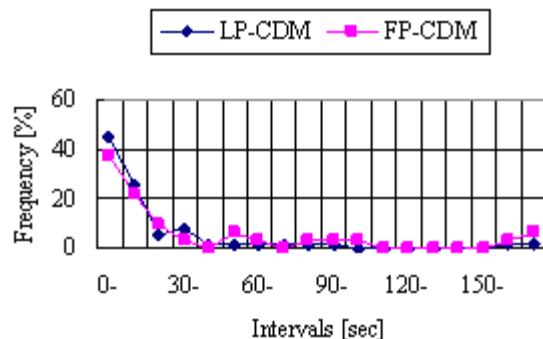


Figure 10. Frequency of CA Intervals

In Fig. 11, pattern ‘↑↓’ is strongly dominant at altitude differences ‘100-’. Conversely, at altitude differences ‘0-’, 3 patterns ‘↑↑’, ‘↓↓’ and ‘→→’ are the majorities. In addition, pattern ‘→→’ rarely occurs.

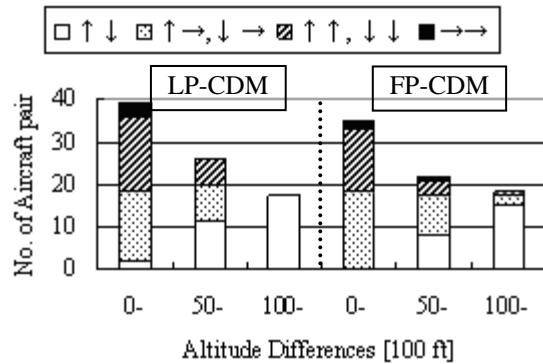


Figure 11. Classification of Vertical Situations

We counted the number of vertical maneuvers by checking the tracks of aircraft pairs at the time of CA occurrence. The result is summarized in TABLE II. The numbers are categorized by altitude differences at the time of CA occurrence. Percentages stand for the numbers divided by the number of aircraft pairs, which are 84 by LP-CDM and 77 by FP-CDM. TABLE II indicates that vertical maneuvers rarely occur at altitude differences ‘100-’ in case of both LP-CDM

and FP-CDM. The majority of vertical maneuvers are considerably connected with the 3 patterns including cruise phase ' \rightarrow '.

TABLE II. NUMBER OF VERTICAL MANEUVER

		0-	50-	100-
LP-CDM	Vertical Maneuvers	16 (19%)	8 (10%)	0 (0%)
	$\uparrow \rightarrow, \downarrow \rightarrow, \rightarrow \rightarrow$	10 (12%)	5 (6%)	0 (0%)
FP-CDM	Vertical Maneuvers	15 (19%)	9 (12%)	1 (1%)
	$\uparrow \rightarrow, \downarrow \rightarrow, \rightarrow \rightarrow$	13 (17%)	5 (6%)	1 (1%)

Considering the results of Fig. 11 and TABLE II, when there is a large altitude difference of more than 10,000 ft, vertical maneuvers are rarely observed. Therefore, CAs which occurred in such situations are regarded as unnecessary and should be suppressed.

VI. EVALUATION OF DAPs-CDM

This section discusses the characteristics of the new DAPs-CDM and describes how the results of simulation verify the effectiveness of DAPs-CDM introduction.

A. Characteristics of DAPs-CDM

The characteristics of DAPs-CDM are explained below by comparing them with the characteristics of conventional CDM. They are to predict aircraft positions using aircraft velocity of DAPs and to determine aircraft flight phases using roll angle and selected altitude of DAPs. TABLE III summarizes the characteristics of DAPs-CDM.

TABLE III. THE CHARACTERISTICS OF DAPs-CDM

	LP-CDM	DAPs-CDM	Remarks
Position $\mathbf{x}_i(t)$	Observed Position by Radar	No Difference	
Velocity $\mathbf{v}_i(t)$	Estimated Velocity	$GS_i(t), TTA_i(t),$ Vertical Speed	Possible to be Optimized for Prediction
Horizontal Flight Phase Determination	Variation of Velocity	Roll Angle	Straight or Turn
Vertical Flight Phase Determination	Assigned Altitude	Selected Altitude	Climb/Descend or Cruise on Prediction Line

The predicted position is calculated by (1) in LP-CDM and the prediction course of aircraft tracks is bent by depending on flight plan information in the similar way like FP-CDM. Where DAPs-CDM differs from the conventional CDM is in using

aircraft velocity (ground speed, true track angle, vertical rate) of DAPs instead of velocity in radar data. This aircraft velocity can be optimized for prediction by smoothing.

When determining aircraft flight phases, the absolute value of roll angle on the airborne side is newly used in horizontal detection. Selected altitude on the airborne side instead of assigned altitude on the ground side is also used in vertical detection. Roll angle is sometimes called bank angle because it stands for the amount of aircraft banking. Selected altitude stands for the control target of altitude.

DAPs can reflect the latest, more accurate state and intent of aircraft. Therefore, DAPs-CDM could reduce prediction errors in position caused by fluctuations in aircraft velocity. In addition, it could predict the change of aircraft flight phases earlier and aircraft positions closer to actual aircraft trajectories.

B. Evaluation of DAPs-CDM Introduction

For the purpose of comparing DAPs-CDM with conventional CDM, ENRI developed CDES (Conflict Detection Evaluation System). It can simulate both DAPs-CDM and conventional CDM under air traffic situations and system parameters almost the same as it does operational ones.

We evaluated the effectiveness of DAPs-CDM introduction by computer simulation. To simulate an environment where aircraft parameters can be downlinked, flight recorded data including all necessary DAPs were used. Air traffic volume consisted of 2 hours of peak time air traffic. The total number aircraft was 575 and 22 of them had DAPs capability. Calculation timing was every 10 seconds in accordance with the update rate for radar data. The horizontal threshold R_h was set as 5 NM. The vertical threshold R_v was basically set as 700 ft and as 1,600 ft in the case of altitude being over 41,000 ft.

As a result, when comparing the number of CA occurrences using DAPs-CDM with those using conventional CDM, there was a difference of 10 occurrences of CA. In all cases of them, the situation was that one aircraft had DAPs capability and the other didn't. Additionally, the use of selected altitude made large differences in CA occurrence.

TABLE IV. DETAIL OF SIMULATION RESULT

Main Effective Factor in DAPs	No. of CA Difference between DAPs Use/Non-use (Total: 10 pairs)	CA Duration
Selected Altitude	6 pairs	Short/Long
Others	4 pairs	Short

Fig. 12 and Fig. 13 show the effectiveness of DAPs-CDM especially in using selected altitude to determine vertical flight phases. Fig. 12 shows CA occurrence not using DAPs-CDM and Fig. 13 shows it using DAPs-CDM. Red and blue lines represent altitude. Only aircraft represented in red has DAPs capability and its selected altitude is represented by a green line. The time when CA was detected was plotted by circle points on the lines.

In Fig. 13, CA occurrences around flight time 01:45 and 01:50 are suppressed due to using selected altitude. Safety separations could already have been set before 01:45 and 01:50. Because selected altitude reflects the latest, more accurate aircraft intent, it is very effective to leverage it proactively, e.g. when checking the agreement with assigned altitude on the ground side.

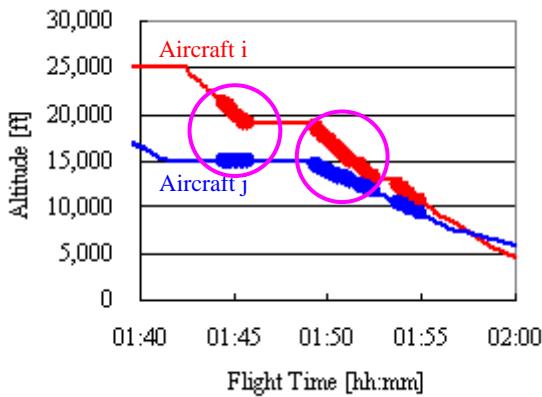


Figure 12. CA Occurrence of Non-Use of Selected Altitude

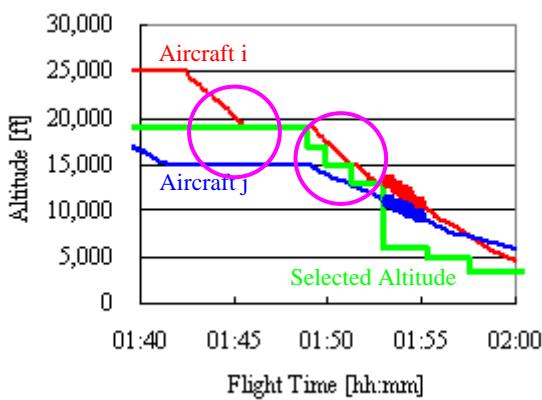


Figure 13. CA Occurrence Using Selected Altitude

VII. CONCLUSION

The aim of this study is to develop a new DAPs-CDM and evaluate the effectiveness of its introduction by computer simulation.

Firstly, in order to understand the characteristics of the conventional CDM, we calculated horizontal and vertical prediction errors in aircraft position. Both horizontal and vertical prediction errors using flight recorded data were reduced in comparison with those using radar data. Horizontal prediction error was reduced by 37% on average in straight phase and reduced by 28% on average in turning phase. Standard deviation of vertical prediction error was reduced by

23%. It was found better to smooth vertical speed for prediction and to use selected altitude in DAPs-CDM.

We then analyzed CA occurrences by conventional CDM. Intermittent CA might often occur in a conventional CDM. When there was a large altitude difference, of more than 10,000 ft, vertical maneuvers were rarely observed. Therefore, CAs which occurred in such situations were regarded as unnecessary and were suppressed.

Finally, the characteristics of DAPs-CDM were studied and the results of simulation which verify the effectiveness of its introduction were demonstrated. The characteristics of DAPs-CDM are to predict aircraft positions using aircraft velocity of DAPs and to determine aircraft flight phases using roll angle and selected altitude of DAPs. As a result, when comparing the number of CA occurrences using DAPs-CDM with those using the conventional CDM, there was a difference of 10 occurrences of CA. We found that the use of selected altitude made large differences in CA occurrence.

ACKNOWLEDGMENT

We are very grateful to all those concerned with this study.

REFERENCES

- [1] ATCA-J (Air Traffic Control Association - Japan), Introduction of Air Traffic Control, 1999.
- [2] NTT DATA Corporation, and JCAPF (Japan Civil Aviation Promotion Foundation), RDP System Overview, 1997.
- [3] JCAB (Japan Civil Aviation Bureau), and JCAPF, Basic Research Reports on International Trend of SSR Mode S in FY 2002, 2002.
- [4] A. Senoguchi, and Y. Fukuda, "Study on Conflict Detection Method with Downlink Aircraft Parameters," Electronic Navigation Research Papers, No. 122, pp.39-51, 2009.
- [5] ICAO, "Procedures for Air Navigation Services - Air Traffic Management," Doc. 4444 ATM/501, pp. 15 - (12-13), 2005.
- [6] EUROCONTROL, "STCA & ACAS Interaction and Interoperability Workshop Report," pp. 11, 2007.
- [7] EUROCONTROL, Specification for Short Term Conflict Alert, 2006.
- [8] EUROCONTROL, Guidance Material for Short Term Conflict Alert Appendix A: Reference STCA System, 2006.
- [9] Council of European Union, European Air Traffic Management Master Plan, edition 1 -30, 2009.
- [10] ICAO, Annex 10 to the Convention on International Civil Aviation Volume IV (Surveillance Radar and Collision Avoidance Systems), 1998.
- [11] EUROCONTROL, Mode S Enhanced Surveillance - 3 States Project - Master Plan, 2002.
- [12] T. Koga, "SSR Mode S related researches in ENRI," EIWAC (ENRI International Workshop on ATM / CNS) 2009, pp. 109-114, Mar. 2009.
- [13] Ministry of Transport, White Book of Transport, Vol. III, Chapter 3, Section 1, 1980.
- [14] NTT DATA Corporation, and JCAPF, "Grafical Leaflet of RDP," pp.202, 1997
- [15] G. V. Morris, Airborne Pulsed Doppler Radar, Artech House, Norwood, 1988.
- [16] A. Senoguchi, and Y. Fukuda "An Analysis on Intermittent Conflict Alert for Air Traffic Control," Proceedings of 2008 KSAS-JSASS Joint International Symposium on Aerospace Engineering, pp. 518-521, 2008.

En Route Air Traffic Control Input Devices for the Next Generation

Matthew J. Mainini

San Jose State University Research Foundation
 & NASA Ames Research Center
 Moffett Field, CA, United States
 matthew.j.mainini@nasa.gov

Abstract—The purpose of this study was to investigate the usefulness of different input device configurations when trial planning new routes for aircraft in an advanced simulation of the en route workstation. The task of trial planning is one of the futuristic tools that is performed by the graphical manipulation of an aircraft's trajectory to reroute the aircraft without voice communication. In this study with two input devices, the FAA's current trackball and a basic optical computer mouse were evaluated with the "pick" button in a click-and-hold state and a click-and-release state while the participant dragged the trial plan line. The trial plan was used for three different conflict types: Aircraft Conflicts, Weather Conflicts, and Aircraft + Weather Conflicts. Speed and accuracy were the primary dependent variables. Results indicate that the mouse conditions were significantly faster than the trackball conditions overall with no significant loss of accuracy. Several performance ratings and preference ratings were analyzed from post-run and post-simulation questionnaires. The release conditions were significantly more useful and likable than the hold conditions. The results suggest that the mouse in the release button state was the fastest and most well liked device configuration for trial planning in the en route workstation.

Keywords-input devices, en route, controller, workstation, mouse, trackball, NextGen

I. INTRODUCTION

A. Background

The United States air traffic control system is expecting such a significant increase in traffic that the current system will not be able to handle it. Controllers have been using the radar scope to maintain separation and efficiency of air travel for more than half a century. Air traffic control has developed considerably over the years, but the fundamentals have remained the same. As traffic continues to increase in the future, the system that we have relied upon for many years may reach a breaking point in which traffic can no longer be safely managed with the current tools and/or operations.

In the en route airspace where aircraft fly at their cruising altitudes, capacity is limited by controller workload, which poses a significant barrier to the projected traffic growth in the future National Airspace System (NAS). For example, domestic figures show that between the years of 2004 and 2005 there was a 6.6 percent increase in the number of flights over

the previous year [1], and overall, the FAA has predicted that by the year 2025, the "total mainline air carrier and regional enplanements are forecast to increase from 757.4 million in 2008 to 1.1 billion in 2025, an average annual rate of 2.2 percent" [1]. To prevent unnecessary accidents and delays, re-evaluating the radar controller workstation is one area that may help reduce the workload of the operator, therefore, aiding the main goal of reducing the strain on the entire system.

B. Input devices

One component of the controller workstation is the input device. Two of the most commonly used input devices when interfacing with a computer are the trackball and the mouse. Compared to the mouse, research has shown that task-based performance is worse when the trackball is used for several tasks including throughput (speed and accuracy), selection, dragging, and tracing [1][5]. Interestingly, one study's results show that participants who used a trackball regularly without ever using a mouse prior to the study performed better with the mouse [5].

C. Button states

Input devices can be operated with various combinations of movements and button presses performed at the same time. The current en route trackball, however, was designed to have the button pressed with no other event to be performed simultaneously. This means that tasks that may have involved selecting items while moving the cursor were not possible and the Display System Replacement (DSR) interface was built with no button press combination features. However, the modern mouse was designed with button state and movement interactions as a primary feature. This would enable several new interactions to exist such as drag-and-drop and selecting items in a group.

There are current day graphical user interfaces (GUIs) that incorporate the use of dragging the device cursor, even hand gestures, with one or more buttons held down to perform a variety of functions that increase the versatility of the device [6]. These combinations allow designers to add different interactions to their applications to effectively make the product easier to use and can decrease the time for completing various tasks. However, previous research shows that holding the button down while moving the cursor increased completion

times for a basic pointing task [3]. Would this effect carry over to the dynamic environment of the DSR with the use of a modern mouse? Are the current input devices capable of additional functionality (e.g. click-and-hold) in the software without degradation of performance?

D. Current study

The Airspace Operations Laboratory (AOL) at NASA Ames Research Center has developed the Multi-Aircraft Control System (MACS) which simulates a wide variety of ATC tools [4]. The specific section of MACS that includes the en route DSR has had many new graphical tools added in recent years that may benefit from a click-hold-and-drag input (e.g. trial planning). The trial plan is a futuristic tool that enables the controller to graphically manipulate the route of an aircraft as he sees fit.

In the present study, the aim is to discover the usefulness of a click-and-drag input device for NextGen via the DSR trial plan task. The trial plan tool is one of the most likely graphical tools to be implemented in future Air Route Traffic Control Center (ARTCC) workstations, and for this reason, it was included in the study. Three types of conflicts were presented to the controllers in which they used the trial plan tool to graphically alter the routes to maintain separation of the aircraft. The hypothesis was that a standard optical computer mouse would result in faster performance than the trackball for the graphical interaction task of trial planning in the DSR. It was also hypothesized that the hold condition would result in faster performance than the release condition when the mouse was used but slower performance when the trackball was used.

II. METHODS

A. Participants

Thirteen participants were selected from a pool of retired local controllers (ages 45-65) from the Oakland control center (ZOA). The number of participants was chosen based on availability. Criteria necessary for participation in the study were extensive experience in air traffic control and prior MACS (Multi-Aircraft Control System) software usage.

B. Design

A 2x2 factorial within-subjects design was used for the study. The independent variables were input device (mouse, trackball) and button state (hold, release). The four conditions in the 2x2 factorial included: Mouse + Hold (MH), Mouse + Release (MR), Trackball + Hold (TH), and Trackball + Release (TR).

The MH condition was performed by pressing and releasing the mouse button on the portal icon (the arrow to the right of the callsign in the first row of the datablock) which opened the manipulatable blue trial plan line. The act of selecting with the left mouse or trackball button is also known as "picking." The controller would then pick on the trial plan line and simultaneously hold it in while dragging. The release of the button dropped the target and completed the mouse action for one trial plan. The controller would then type in the keyboard command "UC CID" (UC = Uplink Clearance, CID

= Computer Identification) to uplink the clearance for the reroute of the selected aircraft. In this condition, picking on the route line and manipulating it was exactly like the Microsoft Windows "drag-and-drop" action. The MR condition was different in that the button was never held, but pressed and released twice instead. The first click grabbed the target while the second click dropped it, completing the mouse action. The TH condition was the same action as the MH condition but with the use of the trackball instead of the mouse. The TR condition was the same action as the MR condition but with the use of the trackball instead of the mouse. The TR condition is currently used as the exclusive device configuration for en route operations.

While manipulating the route of an aircraft, the participants either pressed the button and released it, or pressed the button and held it down. These two conditions are referred to as "release" and "hold," respectively. In the release condition, participants would simply press the button once to grab the trial plan line and press again to drop it. Movement of the device (and cursor) would occur between presses.

The software, MACS, was used in the experiment and was capable of simulating current day operations as well as many possible future concepts. An advanced display containing limited data tags, real-time convective weather, weather probe, and conflict probe was enabled because the hypothesis was constructed with tools that are not yet operational in the real-world (e.g. trial planning). Also, the results likely speak to issues of the future rather than the present.

C. Stimuli

Figure 1 shows the stimuli presented to the controllers. Full datablocks which contained information for a given aircraft including the callsign, altitude, speed, and time-to-conflict (in minutes) appeared in a scripted fashion when the aircraft entered sector ZKC 90 or when a conflict was going to occur within six minutes. All weather-based conflicts appeared when the aircraft crossed the sector boundary and all aircraft conflicts initially appeared at the six minute mark. Six minutes was chosen to provoke the controller to act immediately while still providing a large enough buffer if they needed more time to resolve the conflict or if they were behind due to working on other conflict resolutions.

The conflicts were determined by an algorithm that efficiently predicted if the aircraft would be in conflict with weather or another aircraft. When a conflict appeared, a salient full datablock popped up on the screen to call attention to it. The controller then picked on the portal (arrow to the right of the callsign) to open the trial plan line in which they were able to manipulate.

The controllers' sector of responsibility was ZKC 90, which is a real-world sector inside of Kansas City Center's airspace. Each participant was in a standalone configuration in which no networking was needed between machines. All workstations presented the same scenario simultaneously.

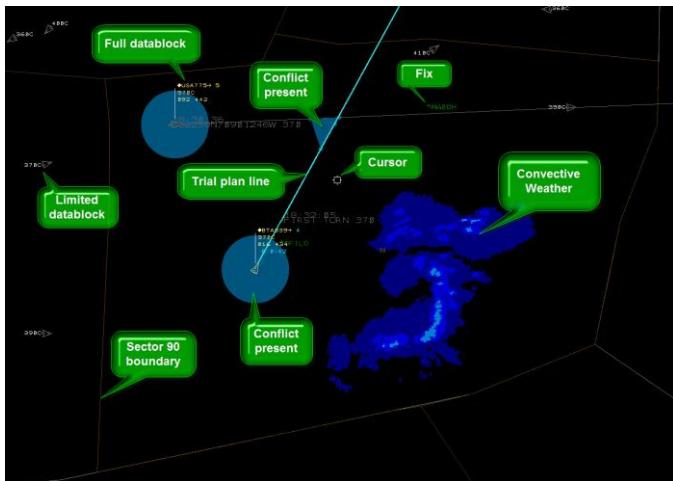


Figure 1. Stimuli presented to the participants on the Display System Replacement.

A convective weather cell was located in the southeast corner of the sector that closed off about 25% of the sector from use. The fixes “MABOH” and “OFLIO” were chosen based on their location in reference to the weather cell. Convective weather was present for the entire duration of the trial and slowly moved east (~5-10kts).

Aircraft were all “owned” by the participant and required no check-ins or handoffs (assumed to be automated). Limited data tags were used because it was an advanced concept in which the controllers were monitoring for conflicts rather than actively solving them. It also served as the primary goal for the participants to maintain all limited datablocks as often as possible.

The participants were instructed not to use any alternatives to the lateral route maneuver such as radio communication, vertical maneuver, slowing the aircraft, or keyboard input. This forced the participant to query a trial plan for an aircraft in conflict. Trial planning is the graphical manipulation of an aircraft’s 4D trajectory. When the portal was opened, a blue line appeared on top of the filed flight path that was extended from the nose of the aircraft to the destination airport, typically with several waypoints along the route. The blue line was then manually picked on and dragged to a new location and dropped to lock it in place. The controller was then able to uplink the clearance to confirm the new route for the aircraft. This reroute process can be done with no voice communication; therefore, voice communication was not necessary for rerouting aircraft in the study.

D. Types of conflicts

To maximize the number of conflicts presented while reducing redundancy and the learning effect, three types of conflicts were scripted within different areas of the sector. The most basic conflicts involved one aircraft and weather. When an aircraft in conflict with weather entered the sector, a full datablock appeared automatically with a blue number indicating how many minutes were left until the aircraft would enter the weather (Figure 2). The full datablock’s appearance was their cue to resolve the conflict. When a full datablock appeared, the steps to complete the task were to pick on the

portal, reroute the aircraft graphically around the weather cell over a specified fix, drop the route on the fix, and finally uplink the clearance via keyboard command. All aircraft that were headed west in conflict with weather were rerouted over the fix “OFLIO,” and aircraft headed east were rerouted over the fix “MABOH.” The controllers were to be as precise as possible while still solving the conflicts quickly. They were told to “put the cross in the box,” which represented the fix and cursor, respectively. A total of ten conflicts of this type were included in each trial.

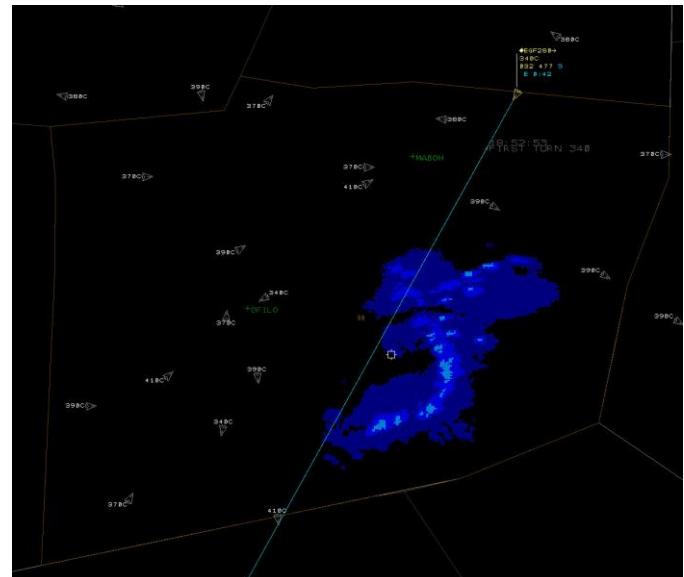


Figure 2. Aircraft flying into weather unless acted upon by a controller.

The second conflict type was between two aircraft. Unlike the weather conflicts, the controllers had some flexibility in the location of the reroute.

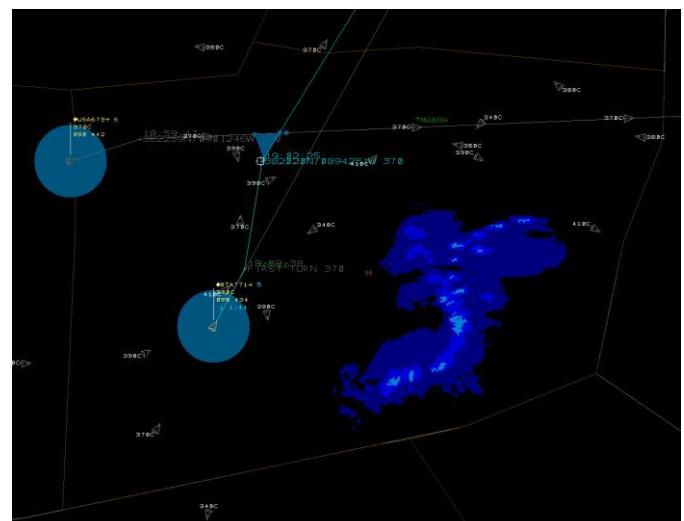


Figure 3. Two aircraft in conflict while the new route (blue line) was moved to the left.

For consistency the controllers were also instructed to maneuver the southernmost aircraft “behind” the other aircraft in conflict (Figure 3). Manipulation of the route so that the rerouted aircraft flew behind the other was typically a more reliable method of conflict avoidance and thus should have been the controllers’ default response regardless of instruction. Real-time feedback for successful conflict avoidance was supported by the disappearance of the large blue circles that indicated a conflict was present as the controller manipulated the trial plan line. A total of ten conflicts of this type were included in each trial.

The third and final conflict type was a combination of a weather conflict followed by an aircraft conflict. This task first involved a reroute around weather over MABOH, exactly like the first conflict type for eastbound aircraft. When the controller rerouted the aircraft over MABOH to avoid the weather, a second conflict would appear with another aircraft along the new route (Figure 4). The controller then moved the line once more to manipulate the aircraft to safely fly behind the other one in conflict while leaving the initial reroute over MABOH alone. This made the conflict resolution more difficult and longer than the other types. The controllers would normally attempt to locate a single fix to resolve both weather and aircraft conflicts in the real-world if possible, however, the instructions were necessary so that each controller resolved the conflicts in the same manner. A total of five conflicts of this type were included in each trial.

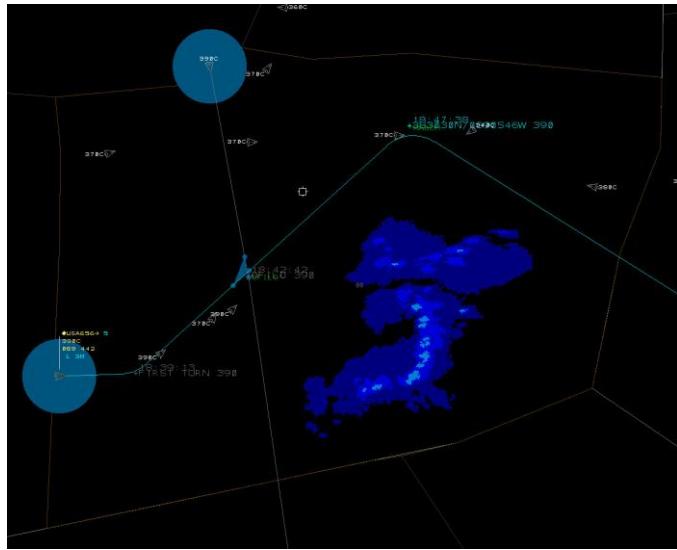


Figure 4. Two aircraft in conflict after one was rerouted around weather.

E. Dependent variables

Objective metrics included in the analysis were time and accuracy. The time to completion of the trial plan was measured as the initial pick on the portal to the final pick that dropped the trial plan line to lock the route in place. Between those picks was when the controller actually dragged the line to the new location, therefore, only the opening of the trial plan line to the release of the line was considered part of the trial plan time.

The accuracy of the pick on the waypoints was measured as the distance in nautical miles from the waypoint. The waypoints had a cross (+) to indicate the exact location of the fix. The cross was the aiming point for the participants to drop the trial plan line. MACS automatically recorded the point (x, y) that the participants actually dropped the line. The point was then compared to the known location of the waypoint.

Subjective metrics were analyzed from questionnaires taken by the participants after each trial and again at the end of the simulation.

III. RESULTS

A repeated measures ANOVA was conducted for all of the results that follow. Table 1 summarizes the F-values and p-values for each of the analyses with bold values to indicate significance ($\alpha = 0.05$). The factors (Device = D, Button State = BS) are listed in the top cells of the columns and the dependent variables are listed in the first column with their respective F-values and p-values to the right. Interactions that were found to be significant were followed up with a post hoc paired samples T-test with the Bonferroni correction.

Table 1. Summary of repeated measures ANOVA results.

Factor > Value >	D F	D p	BS F	BS p	DxBS F	DxBS p
Objective Metrics						
AC Conflict	5.13	0.043	0.29	0.603	5.22	0.041
WX Conflict	74.86	0.000	19.69	0.001	1.24	0.287
AC – WX Conflict	24.87	0.000	0.55	0.474	1.43	0.255
Accuracy	0.88	0.369	0.06	0.805	0.06	0.818
Subjective Ratings						
Workload	0.32	0.584	1.77	0.209	0.32	0.584
Usability	10.55	0.007	5.67	0.035	3.42	0.089
Usefulness	5.15	0.043	31.13	0.000	2.54	0.137
Accuracy of picking	2.25	0.168	1.16	0.309	4.97	0.053
Accuracy moving TP	7.36	0.024	9.53	0.013	6.00	0.037
Cursor speed	19.31	0.001	8.67	0.012	15.60	0.002
TP satisfaction	4.52	0.055	7.02	0.021	1.68	0.219
Comfort level	17.91	0.001	11.93	0.005	12.91	0.004
Likability	12.54	0.004	65.61	0.000	7.92	0.016

A. Trial plan completion time

The main goal and only task for the participants was to successfully trial plan aircraft to avoid conflicts with convective weather and/or other aircraft. The time to completion of the trial plan began with the opening of the portal and ended when the route line was dropped. These events were recorded with MACS internal data logging as well as video screen captures. The keyboard command to uplink the

clearance to finalize the process was not included in the analysis as keyboard inputs would have introduced unnecessary variables.

The metrics for the time to complete a single trial plan were measured in milliseconds but have been rounded for this paper. The three conflict types were calculated separately to show the results individually by conflict type, however, they cannot be directly compared as the trial plan manipulation was not the same for each conflict (i.e. the weather conflicts had specified waypoints to reroute to, while the aircraft conflicts had no specified points and minimizing delay was a priority).

1) Aircraft Conflict

Aircraft conflicts were resolved by rerouting the southernmost aircraft behind the other. The controllers were instructed to quickly resolve the conflicts while minimizing delay. The descriptive statistics for the conditions are MH ($M = 7.87$, $SD = 1.98$), MR ($M = 8.66$, $SD = 3.53$), TH ($M = 10.49$, $SD = 3.33$), TR ($M = 9.15$, $SD = 2.52$). The results of a repeated measures ANOVA show a significant main effect was found for device type ($F_{(1,12)} = 5.13$, $p < 0.05$) and interaction effect ($F_{(1,12)} = 5.22$, $p < 0.05$). The mouse was significantly faster than the trackball when trial planning for conflict avoidance between two aircraft, as shown in Figure 5.

A paired samples T-test with the Bonferroni correction was calculated to find which pair of means were significantly different. The interaction suggested that the MH took less time than the MR condition and the TH condition took longer than the TR condition. However, of the six possible combinations, only the MH condition's trial plan was significantly faster than the TH condition [MH-TH ($t(12) = 3.15$, $p < 0.0083$)].

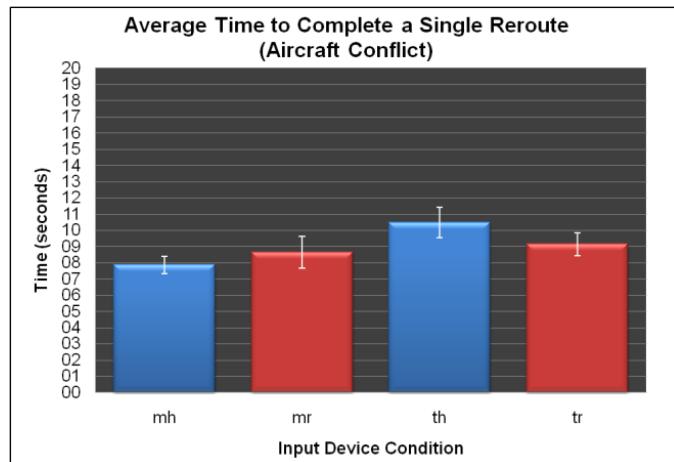


Figure 5. The average time to complete a single reroute for an aircraft conflict.

2) Weather Conflict

Weather conflicts were resolved by rerouting the aircraft around the weather over a specified waypoint (i.e. MABOH, OFILO) depending on the heading of the aircraft. The descriptive statistics for the conditions are MH ($M = 7.65$, $SD = 2.38$), MR ($M = 6.89$, $SD = 1.66$), TH ($M = 10.54$, $SD = 2.32$), TR ($M = 8.83$, $SD = 2.59$). The results of a repeated measures ANOVA show a significant main effect was found

for device type ($F_{(1,12)} = 74.86$, $p < 0.001$) and button state ($F_{(1,12)} = 19.69$, $p < 0.01$). Unlike the Aircraft Conflict results, there was no significant interaction ($F_{(1,12)} = 1.24$, $p > 0.05$). Similar to the Aircraft Conflict results, the mouse was significantly faster than the trackball for trial planning around weather to a specified waypoint. In addition, the release conditions were significantly faster than the hold conditions (Figure 6).

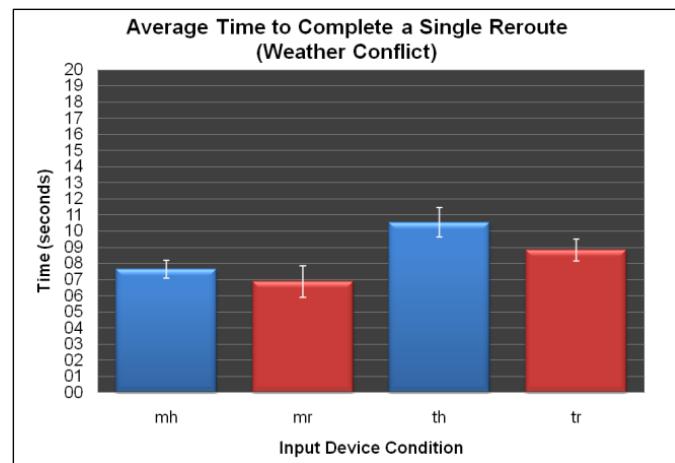


Figure 6. The average time to complete a single reroute for a weather conflict.

3) Aircraft + Weather Conflict

The aircraft + weather conflicts involved a combination of the two conflict types during a single trial plan. This added more complexity while lengthening the duration of the trial plan. The descriptive statistics for the conditions are MH ($M = 12.42$, $SD = 4.34$), MR ($M = 12.77$, $SD = 2.93$), TH ($M = 17.50$, $SD = 4.21$), TR ($M = 16.18$, $SD = 4.06$). The results of a repeated measures ANOVA show a significant main effect was found for device type ($F_{(1,12)} = 24.87$, $p < 0.001$). There was no main effect for button state ($F_{(1,12)} = 0.55$, $p > 0.05$). For this conflict type, the mouse was much faster than the trackball (Figure 7). Although Figure 7 suggests an interaction between device type and button state, the interaction was not significant ($F_{(1,12)} = 1.43$, $p > 0.05$).

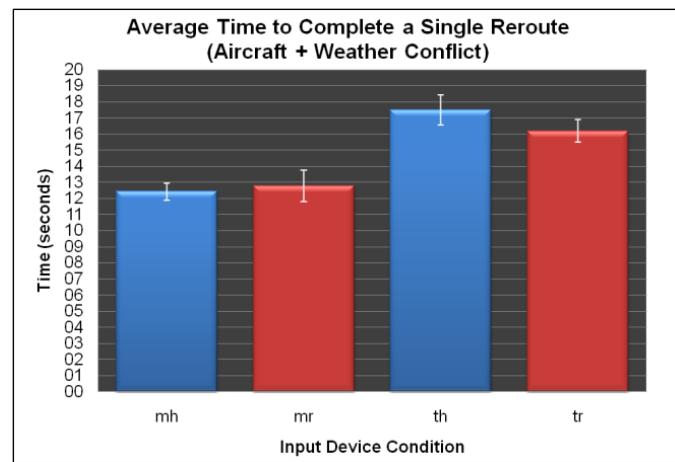


Figure 7. The average time to complete a single reroute for an aircraft + weather conflict.

B. Accuracy of picking on the waypoints

Accuracy was measured by the proximity of the cursor to a waypoint that the participants were instructed to reroute the aircraft over (for conflicts that involved weather). They were also instructed to resolve the conflicts quickly while dropping the trial plan line as accurately as possible on the waypoint. Due to an inherent tradeoff between speed and accuracy, the precision of the pick on the waypoints MABOH and OFILO were analyzed with a repeated measures ANOVA [MH ($M = 0.33$, $SD = 0.09$), MR ($M = 0.33$, $SD = 0.10$), TH ($M = 0.31$, $SD = 0.15$), TR ($M = 0.30$, $SD = 0.08$)]. No statistical significance was found in the analysis. The overall precision of all four conditions was between 0.30 to 0.35 nautical miles with very little variance (0.01). Overall, there was no speed/accuracy tradeoff.

C. Workload

The workload level was immediately rated after completing a trial on a scale from 1 to 5 (Very low (1) - Very high (5)). Workload was defined as the overall cognitive demand on the participant while resolving conflicts. The hold conditions [MH ($M = 1.31$, $SD = 0.48$), TH ($M = 1.38$, $SD = 0.65$)] were rated as having slightly more of a workload than the release conditions [MR ($M = 1.15$, $SD = 0.38$), TR ($M = 1.15$, $SD = 0.38$)]. Workload ratings were very low for all conditions and the repeated measures ANOVA analysis showed no significance for main effects or interaction effect.

D. Participant ratings

Subsequent results were analyzed from participant ratings on post-run and post-simulation questionnaires. A Likert scale from 1 to 5 was used to capture the controllers' perception about specific aspects of the device configurations.

1) Usability ratings

Participants rated the usability (ease of use) of each device configuration on a 1 to 5 scale (Not easy (1) - Very easy (5)), [MH ($M = 4.69$, $SD = 0.63$), MR ($M = 4.84$, $SD = 0.38$), TH ($M = 3.85$, $SD = 0.99$), TR ($M = 4.54$, $SD = 0.52$)]. Results from the repeated measures ANOVA show a significant effect for device type ($F_{(1,12)} = 10.55$, $p < 0.01$) and button state ($F_{(1,12)} = 5.67$, $p < 0.05$). Regardless of the button state, the mouse was rated as significantly more usable than the trackball. Also, the release conditions were rated as significantly easier to use than the hold conditions (Figure 8).

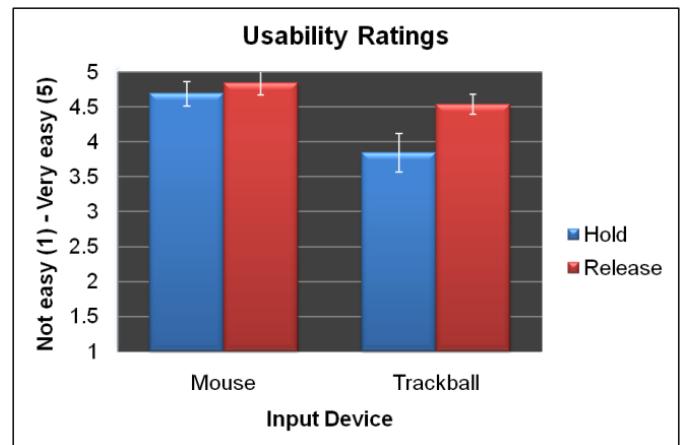


Figure 8. The usability of the device conditions as rated by the participants after each trial.

2) Usefulness ratings

The usefulness of the device configurations was rated on a 1 to 5 scale (Not useful (1) - Very useful (5)), [MH ($M = 4.23$, $SD = 0.83$), MR ($M = 4.85$, $SD = 0.38$), TH ($M = 3.62$, $SD = 0.77$), TR ($M = 4.62$, $SD = 0.51$)]. The participants' rating of the device type show that the mouse was significantly more useful than the trackball ($F_{(1,12)} = 5.15$, $p < .05$). The button state release condition was also significantly more useful than the hold condition ($F_{(1,12)} = 31.13$, $p < .01$). Graphically, the chart looks similar to the usability ratings (Figure 8) with the button state difference more pronounced (i.e. the hold condition values are lower).

3) Accuracy of pick action ratings

The accuracy of the pick action was rated on a 1 to 5 scale (Not accurate (1) - Very accurate (5)), [MH ($M = 4.73$, $SD = 0.47$), MR ($M = 4.42$, $SD = 0.51$), TH ($M = 4.15$, $SD = 0.8$), TR ($M = 4.64$, $SD = 0.50$)]. The repeated measures ANOVA analysis showed no significance for main effects or interaction effect.

4) Accuracy of moving trial plan line ratings

The accuracy of moving the trial plan was rated on a 1 to 5 scale (Not accurate (1) - Very accurate (5)), [MH ($M = 4.73$, $SD = 0.47$), MR ($M = 4.75$, $SD = 0.45$), TH ($M = 3.54$, $SD = 1.13$), TR ($M = 4.73$, $SD = 0.47$)]. The results show a significant effect for device type ($F_{(1,12)} = 7.36$, $p < .05$), button state ($F_{(1,12)} = 9.53$, $p < .05$), and interaction effect ($F_{(1,12)} = 6.00$, $p < .05$).

A paired samples T-test with the Bonferroni correction was calculated to find which pair of means were significantly different. Of the six possible combinations, the three that included the TH condition were significant [MH-TH ($t(10) = 3.46$, $p < 0.0083$), MR-TH ($t(11) = 4.10$, $p < 0.0083$), TH-TR ($t(10) = 3.36$, $p < 0.0083$)]. The TH condition was rated significantly lower than the other three conditions which were rated as very accurate (Figure 9).

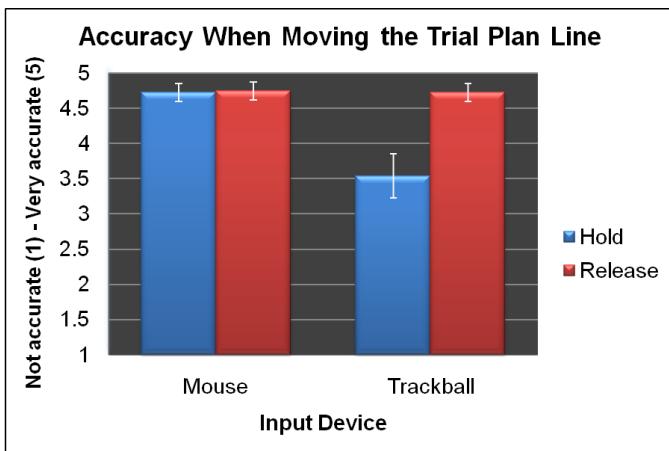


Figure 9. Accuracy of moving trial plan line ratings.

5) Speed of cursor movement ratings

The speed of the cursor was rated on a 1 to 5 scale (Not fast (1) - Very fast (5)), [MH ($M = 4.77$, $SD = 0.44$), MR ($M = 4.77$, $SD = 0.44$), TH ($M = 3.46$, $SD = 0.88$), TR ($M = 4.46$, $SD = 0.52$)]. Cursor movements occurred when the participant manipulated the device to move the cursor around on the screen whether or not the button was held down. A significant result was found for device type ($F_{(1,12)} = 19.31$, $p < .01$), button state ($F_{(1,12)} = 8.67$, $p < .05$), and interaction effect ($F_{(1,12)} = 15.60$, $p < .01$).

A paired samples T-test with the Bonferroni correction was calculated to find which pair of means were significantly different. Of the six possible combinations, the three that included the TH condition were significant [MH-TH ($t(12) = 4.98$, $p < 0.0083$), MR-TH ($t(12) = 4.57$, $p < 0.0083$), TH-TR ($t(12) = 3.61$, $p < 0.0083$)]. The TH condition's cursor movement speed ratings were significantly slower than the other three conditions. Graphically, the chart resembles that of the Accuracy When Moving the Trial Plan Line (Figure 9) with the TR value slightly lower.

6) Trial planning satisfaction ratings

Trial planning satisfaction was used to gauge which conditions the participants felt a sense of satisfaction when performing [MH ($M = 4.54$, $SD = 0.66$), MR ($M = 4.85$, $SD = 0.38$), TH ($M = 4.08$, $SD = 0.86$), TR ($M = 4.69$, $SD = 0.48$)]. Ratings were on a 1 to 5 scale (Not satisfying (1) - Very satisfying (5)). The trial planning satisfaction results (Figure 10) are similar to the usability and usefulness results. The button state release condition was significantly more satisfying to trial plan than the hold condition ($F_{(1,12)} = 7.02$, $p < .05$). The device type and interaction effect were not significant.

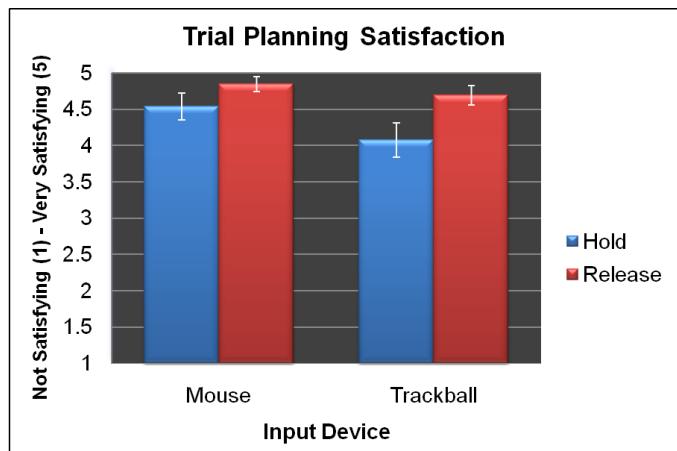


Figure 10. Satisfaction of trial planning ratings.

7) Comfort level ratings

The overall comfort of using the device configurations was rated on a 1 to 5 scale [MH ($M = 4.85$, $SD = 0.38$), MR ($M = 4.92$, $SD = 0.28$), TH ($M = 3.31$, $SD = 1.32$), TR ($M = 4.62$, $SD = 0.51$)]. A significant result was found for device type ($F_{(1,12)} = 17.91$, $p < .01$), button state ($F_{(1,12)} = 11.93$, $p < .01$), and interaction effect ($F_{(1,12)} = 12.91$, $p < .01$).

A paired samples T-test with the Bonferroni correction was calculated to find which pair of means were significantly different. Of the six possible combinations, the three that included the TH condition were significant [MH-TH ($t(12) = 4.38$, $p < 0.0083$), MR-TH ($t(12) = 4.40$, $p < 0.0083$), TH-TR ($t(12) = 3.58$, $p < 0.0083$)]. The overall comfort level of the TH was significantly less than the other conditions (Figure 11). Comfort level when picking and when moving the trial plan line were also gathered. The results mirror that of the overall comfort level.

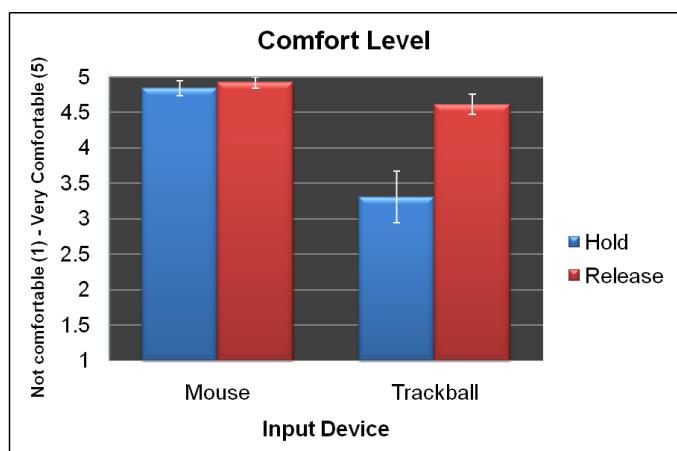


Figure 11. Comfort level of device configuration ratings.

8) Likability ratings

A post-simulation questionnaire was completed by the participants after they experienced each of the four conditions. The likability ratings greatly favored the release conditions

[MH ($M = 3.73$, $SD = 0.44$), MR ($M = 4.77$, $SD = 0.44$), TH ($M = 2.46$, $SD = 1.05$), TR ($M = 4.54$, $SD = 0.66$)]. The results were significant for device type ($F_{(1,12)} = 12.54$, $p < .01$), button state ($F_{(1,12)} = 65.61$, $p < .01$), and interaction effect ($F_{(1,12)} = 7.92$, $p < .05$). The mouse was significantly more likable than the trackball and the release button state was significantly more likable than the hold button state. A paired samples T-test with the Bonferroni correction was calculated to find which pair of means were significantly different. Of the six possible combinations, five were significant [MH-MR ($t(12) = 11.69$, $p < 0.0083$), MH-TH ($t(12) = 3.93$, $p < 0.0083$), MH-TR ($t(12) = 3.41$, $p < 0.0083$), MR-TH ($t(12) = 7.04$, $p < 0.0083$), TH-TR ($t(12) = 5.67$, $p < 0.0083$)]. The release conditions were very well liked and were not significantly different (Figure 12).

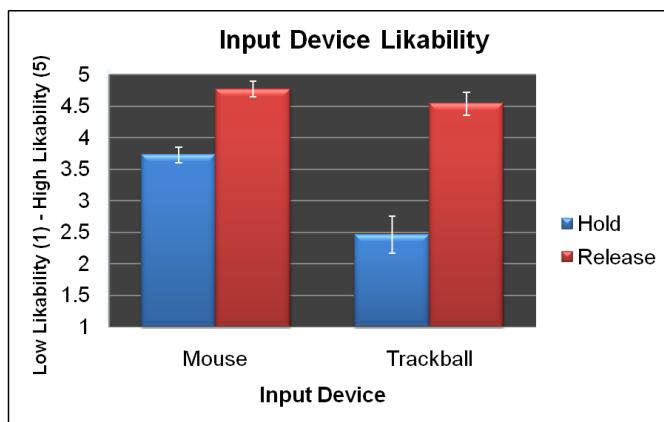


Figure 12. Device configuration likability ratings.

IV. DISCUSSION

The purpose of this study was to investigate the usefulness of different input device configurations when trial planning new routes for aircraft in an advanced simulation. The four conditions were created to discover the advantages and/or disadvantages of each device configuration.

The results partially support the hypotheses because the mouse was significantly faster than the trackball for all conflict types. However, the MH and MR conditions were not significantly different for Aircraft Conflicts and Aircraft + Weather Conflicts. In fact, the MH condition was significantly slower than the MR condition for weather conflicts possibly due to the controllers' familiarity with the release condition.

There was no significant loss in accuracy even though the mouse conditions were performed significantly faster than the trackball conditions. This finding suggests that the mouse had the same accuracy strength but was much faster than the trackball. It appears that the mouse was a superior device for the trial plan task regardless of button state; however, the release condition was liked significantly more. If implemented in future workstations, controllers may enjoy the option to choose between the MR and MH conditions.

The mouse conditions were rated as significantly more usable overall, when picking, and when trial planning. The

MR condition was considered the most usable device configuration, but the performance was not significantly different from the MH condition for the Aircraft Conflict and Aircraft + Weather Conflict types. However, the MH condition has the advantage of additional interactions that may benefit the future en route workstation depending on the toolset required.

Interestingly, The TR condition was rated as more likable than the MH condition even though performance was worse for the TR condition. The preference may be due to the extensive use with the TR over years of working traffic with that configuration. It was also present in the trial plan satisfaction ratings and the usefulness ratings, likely due to the general familiarity of the release condition. Issues such as reliability and trust may have been some of the underlying factors that influenced the ratings.

As expected, the TH condition was the slowest and least liked for trial planning in every case. The trackball was difficult to use when the participants were forced to hold the pick button and move the cursor with one hand.

ACKNOWLEDGMENT

This research was supported by San Jose State University Research Foundation and the Airspace Operations Laboratory at NASA Ames Research Center in Mountain View, California, United States.

REFERENCES

- [1] Federal Aviation Administration (2009). FAA forecast fact sheet. Retrieved October 19, 2009 from the Federal Aviation Administration website: http://www.faa.gov/news/fact_sheets/news_story.cfm?newsId=10457.
- [2] MacKenzie, S., Kauppinen, T., Silfverberg, M. (2001). Accuracy measures for evaluating computer pointing devices. CHI 2001. Vol. 3-1.
- [3] MacKenzie, S., Sellen, A., Buxton, W. (1991). A Comparison of input devices in elemental pointing and dragging tasks. Proceedings of the SIGCHI conference on Human Factors in computing systems: Reaching through technology. p. 161-166.
- [4] Prevot, T., Smith, N., Palmer, E., et. al. (2006). The Airspace Operations Laboratory (AOL) at NASA Ames Research Center. Proceedings of the AIAA Modeling and Simulation Technologies Conference and Exhibit, 2006, Keystone, Colorado.
- [5] Sperling, B.B., Tullis, T.S. (1988). Are you a better "mouser" or "trackballer"? A comparison of cursor-positioning performance. SIGCHI Bulletin. Vol. 19-3.
- [6] Yang, J., Choi, E., Chang, W. (2004). A novel hand gesture input device based on inertial sensing technique. Proceedings of the Industrial Electronics Society, 2004, IECON 2004, 30th Annual Conference of IEEE, Vol. 3, p. 2786-2791.

Track 5

Decision Support Tools

An Advanced Particle Filtering Algorithm for Improving Conflict Detection in Air Traffic Control

Ioannis Lymeropoulos
 Automatic Control Laboratory
 ETH Zurich
 Zurich, CH-8092, Switzerland
 Email: lymeropoulos@control.ee.ethz.ch

Georgios Chaloulos
 Automatic Control Laboratory
 ETH Zurich
 Zurich, CH-8092, Switzerland
 Email: chaloulos@control.ee.ethz.ch

John Lygeros
 Automatic Control Laboratory
 ETH Zurich
 Zurich, CH-8092, Switzerland
 Email: lygeros@control.ee.ethz.ch

Abstract—Enhanced accuracy in aircraft conflict detection allows for more efficient use of the airspace and increased safety levels. Trajectory prediction lies at the heart of most conflict detection algorithms. By comparing the predicted trajectories of different aircraft against each other, we can detect real threats while avoiding false alarms. We show how trajectory prediction tools that account for weather forecast errors can improve the performance of a conflict detection scheme. Using information from multiple aircraft at different locations and time instants, wind forecast uncertainties are reduced increasing trajectory prediction accuracy. We present a particle filtering algorithm that can efficiently cope with the high dimensionality and the non-linearity of the problem and show how using this algorithm can improve considerably conflict detection rates in mid and short term horizon encounters.

I. INTRODUCTION

The current Air Traffic Management (ATM) system is to a large extent based on a rigidly structured airspace and a mostly human-operated system architecture [1]. This could potentially impose a constraint in the growth of air traffic, which is otherwise expected to increase considerably the following years [2]. Recent research efforts focus on integrating the segregated airspace (following SESAR [3] in Europe and NEXTGEN [4] in the US). In support of this effort, a large variety of automation and decision support tools are being developed to provide Air Traffic Controllers (ATCs) with more accurate predictive information about aircraft trajectories, local and national traffic flow, weather and routing.

Guaranteing safety in air travel remains the primary concern in the future ATM. One important aspect in this direction is the separation assurance between flight trajectories. Whenever a prescribed minimum separation between two aircraft is violated, a conflict occurs. For conflicts to be identified and prevented, an automated mechanism for Conflict Detection (CD) is required. Once a conflict is predicted, either a centralized [3] or a decentralized [5] Conflict Resolution (CR) scheme can be used to resolve it; for an overview see [6].

CD is itself a challenging task and very often it is combined with the CR task. Most common methods can be divided into three major categories, based on the prediction horizon they consider. Roughly speaking Long term Conflict Detection and Resolution (CD&R) methods deal with horizons of more than 30 mins. Their main concern is typically flow management

problems. Mid term CD&R, accounts for prediction horizons up to 30 mins. Finally, short term CD&R, deals with horizons up to 10 mins.

For the conflict detection to be accurate, one should be able to compute a reliable prediction of the trajectory of an aircraft [7]. Increasing levels of traffic require systems that can accurately predict conflicts earlier, in order to accommodate the extra traffic demand. An automated conflict detection mechanism can take advantage of data that might not be directly accessible, or possibly hard to interpret, by the air traffic controllers, such as the estimated state of the aircraft, weather information and weather uncertainty or different aircraft performance models. This information combined with the data that an air traffic controller has access to, like the estimated position and aircraft flight plans, can lead to an algorithm that improves Trajectory Prediction (TP) and assists the Air Traffic Controller (ATC) in identifying early potential conflicting situations. The longer the horizon the aircraft trajectory is accurately predicted, the more flexibility the ATC (or a conflict resolution algorithm) has to resolve a conflict, or to accommodate more traffic.

Here we demonstrate how CD can be improved by reducing TP inaccuracies related to wind forecast errors. The problem of extracting the wind forecast error information from the trajectories of the aircraft is formulated as a filtering problem. The state that has to be estimated is high dimensional, since it comprises the states (position and heading) of all aircraft in the region of interest, as well as the wind forecast error (projected on a grid). The situation is further complicated by the fact that the aircraft dynamics, through which the wind forecast error is indirectly observed, are nonlinear. This implies that efficient filtering methods (such as the Kalman filter [8]) are inapplicable in this case, whereas methods that could cope with nonlinear dynamics (such as the Particle Filter (PF) [9]) have difficulties dealing with high dimensional states. To solve the problem a novel particle filtering algorithm is developed (called Sequential Conditional Particle Filter (SCPF)) that can deal with both the nonlinear and the high dimensional nature of the problem. For this, the special structure afforded by the filtering problem is exploited, namely the fact that wind forecast error dynamics are linear and conditional on the wind, the dynamics of the different aircraft are independent.

The performance of the proposed algorithm is assessed with a series of simulated feasibility studies. A base scenario involving 2 aircraft flying level with constant airspeed is established. A series of different wind forecast errors is generated which result in potential conflicts. The algorithm shows a significant improvement over the non-filtered wind-forecast especially for the mid-term prediction horizon (20-15 minutes). A further improvement can be achieved when additional aircraft precede the flight of interest.

The rest of the paper is organized as follows. Section II briefly describes the aircraft and wind models used for the simulations. Section III introduces general nonlinear filtering and Particle Filters and outlines the proposed algorithm. Finally, Section IV provides simulation results that document the performance of the proposed method. The paper concludes with Section V which states the conclusions of this study and some ideas for future work.

II. MODEL DYNAMICS

We present a Point Mass Model (PMM) that simulates the dynamics of a commercial aircraft from the point of view of an air traffic controller. The model is framed in the context of stochastic hybrid systems and is capable of capturing multiple instances of flights, each with a different flight plan, aircraft dynamics and flight management system. The dynamics capture the effect of the wind and the wind forecast error, which is treated as a stochastic disturbance to the model.

A. Aircraft Dynamics

The model presented here concentrates on level flights with constant airspeed and is a simplified version of a full model (including varying airspeed, altitude, and control of flight path angle) developed in our earlier work [10], [11]. The dynamics of the aircraft are characterized by the following state vector $z = [X, Y, \psi, m] \in \mathbb{R}^4$ where X and Y are the position of the aircraft in the West-East and South-North direction, respectively, m denotes the mass of the aircraft and ψ its heading. We assume that each aircraft flies with known, constant True Airspeed (TAS) which depends on aircraft type and altitude. Figure 1 depicts the major variables of the model. The relation between the states is nonlinear and depends also on the actions of the Flight Management System (FMS). The values of different parameters (for example the TAS, the lift and drag coefficients, or fuel burn coefficient) which depend on aircraft type, the phase of flight and aircraft configuration are obtained from the Base of Aircraft Data (BADA) database [12]. The movement of the aircraft is also affected by the wind which acts as a disturbance. Thus, the equations of motion, for level flight, become

$$\begin{bmatrix} \dot{X}(t) \\ \dot{Y}(t) \\ \dot{\psi}(t) \\ \dot{m}(t) \end{bmatrix} = \begin{bmatrix} V \cos(\psi(t)) + w_X(t) \\ V \sin(\psi(t)) + w_Y(t) \\ \frac{C_L S \rho(Z) V \sin(\phi(t))}{2m(t)} \\ -\eta T(t) \end{bmatrix}. \quad (1)$$

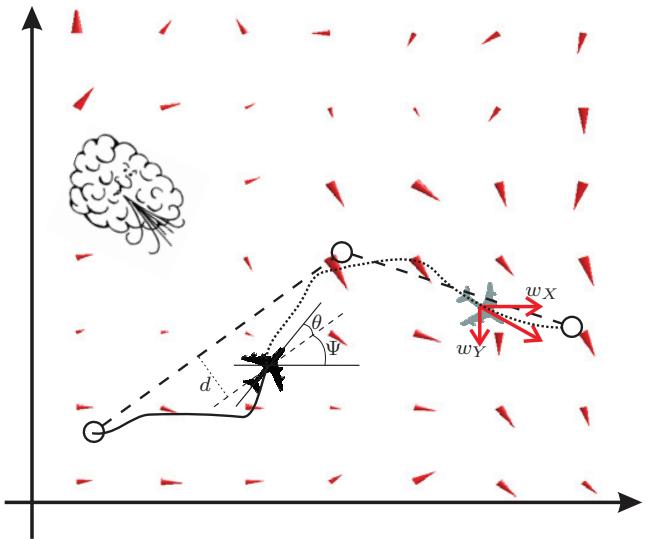


Fig. 1. The aircraft FMS tracks the flight plan between two subsequent way-points, in the presence of wind. The cones represent the direction and magnitude of the wind at different locations. Ψ denotes the nominal heading.

C_L is the lift coefficient, S represents the surface of the wings, $\rho(\cdot)$ the air density depending on altitude (Z). While V is the true airspeed, (ϕ) the bank angle and η the fuel burn coefficient. The values of the parameters (including the TAS) which depend on aircraft type, the phase of flight and aircraft configuration are obtained from the BADA database [12].

B. Flight Management System

The FMS measures the state of the aircraft and guides it along the flight plan by determining the values of the inputs. One of its two components is controlling along track and vertical motion (in our case maintaining constant altitude and airspeed) through the thrust and flight path angle and the other is controlling cross track motion through the bank angle. To ensure constant airspeed the thrust is set equal to the drag force, whereas to ensure level flight we assume that the flight path angle is set to zero. The bank angle is set using a nonlinear feedback controller which corrects cross track deviations from the flight plan encoded through heading (θ) and cross-track errors (d) in Figure 1. Details of the design of these controllers are given in [10], [11]. The aircraft dynamics and the control inputs applied by the FMS are affected by a change in the discrete part of the dynamics. The discrete dynamics arise from the flight plan of the aircraft and the logic variables embedded in the FMS. For more details the reader is referred to [10]–[12].

C. Radar Model

The position of all aircraft is measured using a ground radar. We assume that the radar measurements are corrupted by noise. In practice the accuracy of the radar usually decreases as an aircraft moves away from the radar location. For simplicity, we use the same measurement error statistics for all distances,

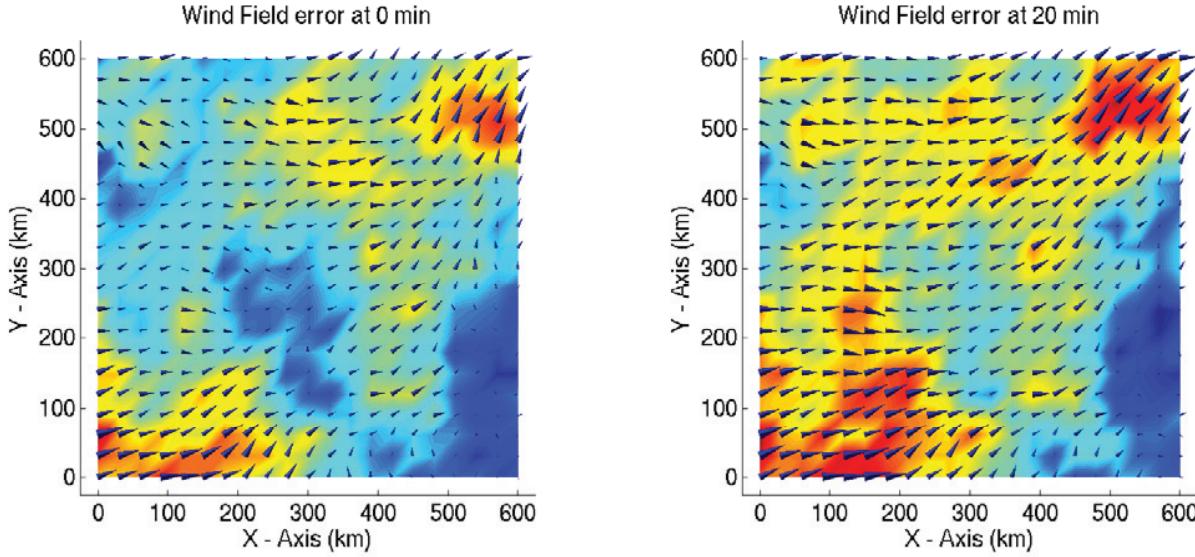


Fig. 2. Example of wind forecast error on a single flight level. The intensity of the color from blue to red indicates a low to high wind forecast error magnitude respectively. The left image displays the initial wind-field, while the right the evolution of the wind-field after 20 minutes. The horizontal grid resolution is 30x30km

and select the variance high enough ($\sigma_r = 80m$) to ensure we err on the side of caution.

D. Wind Model

The uncertainty in the flight trajectory is greatly affected by inaccurate meteorological forecasts [13], [14]. We model the wind as a sum of two components: a nominal component (representing weather forecasts) and a stochastic component (representing forecast errors).

1) Wind Forecast: The nominal part of the wind-field represents the meteorological predictions that are available to the ATC. We obtain those meteorological data from the Rapid Update Cycle (RUC), a numerical weather prediction model for the U.S.A. [15]. The RUC model is run every three hours and each run produces a set of three hourly forecasts.

2) Wind Forecast Error Statistics: We model the wind forecast errors as a random field: $w : \mathbb{R} \times \mathbb{R}^3 \rightarrow \mathbb{R}^2$ where $w(t, P)$ represents the wind at point $P \in \mathbb{R}^3$ and at time $t \in \mathbb{R}$. Since we restrict attention to level flights we ignore wind in the vertical direction. For simplicity, here we restrict attention to the case where $w(t, P) \in \mathbb{R}^2$ is Gaussian with zero mean and covariance matrix $R(t, P, t', P') \in \mathbb{R}^{2 \times 2}$. We calculate the covariance matrix, describing the spatiotemporal correlation of the forecast error based on [16]. The correlation decays exponentially with horizontal distance, altitude and time difference. The data suggest a strong correlation between wind errors in the same horizontal plane, a very strong correlation in time and a weaker correlation across different altitudes.

To describe the wind-field we grid the airspace into a lattice comprising N_X points in the South-North direction, N_Y points in the East-West direction and N_Z points vertically. For each

grid point in the lattice we generate two random numbers, one for the South-North and one for the East-West direction of the wind forecast error. We store these numbers in two vectors $W_X(k)$ and $W_Y(k)$, at time step $k \in \mathbb{N}$ (every δ_t seconds). An example of the horizontal part of the lattice can be seen in Figure 2.

Let $\hat{R} \in \mathbb{R}^{N_X N_Y N_Z \times N_X N_Y N_Z}$ denote the covariance matrix of $W_X(k)$ (by the isotropic assumption, the matrix will be identical for $W_Y(k)$). We generate wind samples using the following linear Gaussian model

$$\begin{aligned} W_X(0) &= \hat{Q}v_X(0), \quad W_X(k+1) = aW_X(k) + Qv_X(k+1), \\ W_Y(0) &= \hat{Q}v_Y(0), \quad W_Y(k+1) = aW_Y(k) + Qv_Y(k+1), \end{aligned} \quad (2)$$

where $v_X(k), v_Y(k) \in \mathbb{R}^{N_X N_Y N_Z}$ are standard (zero mean, identity covariance matrix) independent Gaussian random variables. Q and \hat{Q} are derived by Cholesky Decomposition from the covariance matrix \hat{R} according to

$$QQ^T = (1 - a^2)\hat{R} \text{ and } \hat{Q}\hat{Q}^T = \hat{R}. \quad (3)$$

It is easy to show that the covariance matrices of the resulting vectors for an appropriate choice of a (we set $a = e^{-\delta_t/G_t} \in \mathbb{R}$, where G_t is a parameter of the time correlation [16]) closely resemble the structure implied by the spatiotemporal correlation. Linear interpolation of the wind at the neighboring grid points is used to compute the wind forecast error between the grid points, details can be found in [17].

III. NONLINEAR FILTERING

Problems in engineering applications often require the accurate estimation of the state of a system that evolves in time, using a sequence of noisy observations that become available

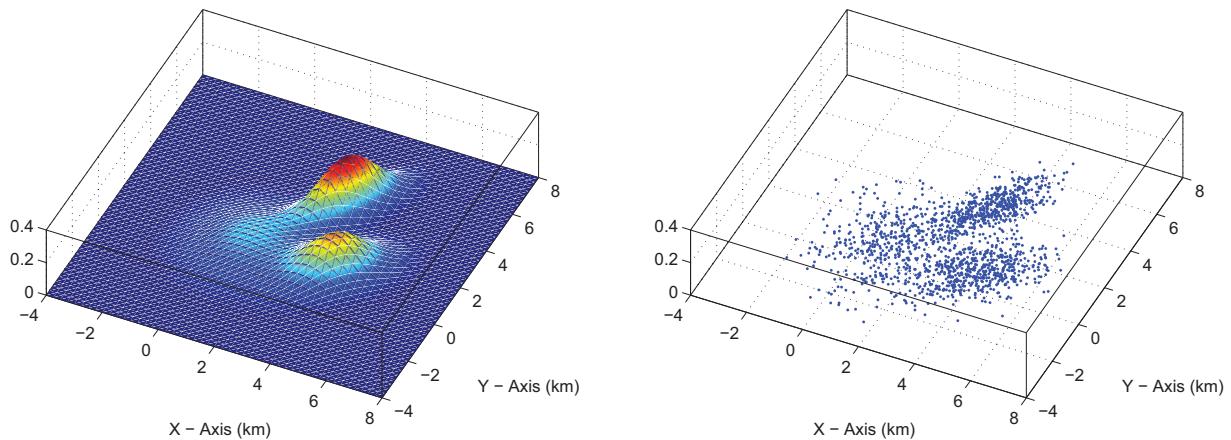


Fig. 3. Example of continuous probability density function (left) and its particle approximation (right). The location and weight of the particles reflect the value of the probability density in that region of the state space.

on-line. For several cases, it is important to include elements of nonlinearity and non-Gaussianity in order to accurately capture the underlying dynamics of the physical system. Moreover, for real time performance it is usually required to process data as they arrive, both from the point of view of storage but mainly for keeping the computational complexity manageable. Filtering algorithms perform such on-line data assimilation to generate estimates of the state [18], [19].

The starting point is typically a discrete time model of the dynamics of the process and the measurements of the form

$$\begin{aligned} x(k+1) &= f(x(k), v(k), k) \\ y(k) &= h(x(k), n(k), k), \end{aligned} \quad (4)$$

where $x(k) \in \mathbb{R}^n$ and $y(k) \in \mathbb{R}^p$ are the state and output of the system at time $k \in \mathbb{N}$, and $f : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{N} \rightarrow \mathbb{R}^n$ and $h : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{N} \rightarrow \mathbb{R}^p$ are (possibly nonlinear) functions. $v(k) \in \mathbb{R}^n$ and $n(k) \in \mathbb{R}^p$ are process and measurement noise, which are generally assumed to be independent, identically distributed stochastic processes, but not necessarily additive, or Gaussian. We also assume that the initial state is independent of the noise processes and its distribution is given through a Probability Density Function (pdf) $p(x(0))$. If the pdf of the noise processes are known, the system of can be equivalently represented using two pdf

$$\begin{aligned} x(k) &\sim p_x(\cdot|x(k-1), k) \\ y(k) &\sim p_y(\cdot|x(k), k). \end{aligned} \quad (5)$$

Here $p_x(\cdot|x(k-1), k)$ is a conditional pdf that models the stochastic dynamics of the state of the system, determined by f and the pdf of $v(k)$, while $p_y(\cdot|x(k), k)$ is a conditional pdf that models the probability distribution of the measurements, determined by h and the pdf of $n(k)$.

Given $k, k' \in \mathbb{N}$ let $\mathbb{Y}(k') = \{y(i)\}_{i=0, \dots, k'}$ denote the sequence of measurements up to time k' and $\mathbb{X}(k) = \{x(i)\}_{i=0, \dots, k}$ denote the sequence of states up to time k . The aim is to estimate the pdf $p(\mathbb{X}(k)|\mathbb{Y}(k'))$. This density

function embodies our best estimate of the state vector up to time k given all available information up to time k' . Depending on the relation of k to k' we can formulate three different types of estimation problems; **Filtering** ($k = k'$), **Prediction** ($k > k'$), **Smoothing** ($k < k'$). These can be solved recursively by invoking Bayes' theorem.

A. Particle Filters

The analytical solution of the optimal Bayesian estimate is not always possible, since the integrals involved are seldomly tractable. In the general case we need to approximate numerically the pdf of interest. Particle filters (or Sequential Monte Carlo methods [9]) are fast estimation techniques that perform this numerical approximation using simulation. The main idea is to approximate the continuous probability distribution of interest using a discrete distribution comprising weighted samples (known as particles, Figure 3). To do this we extract N independent identically distributed particles, $\mathbb{X}^1(k), \dots, \mathbb{X}^N(k)$ from $p(\mathbb{X}(k)|\mathbb{Y}(k))$, and construct an empirical estimate of the distribution

$$\hat{p}(\mathbb{X}(k)|\mathbb{Y}(k)) = \frac{1}{N} \sum_{i=1}^N \delta_{\mathbb{X}^i(k)}(\mathbb{X}(k)), \quad (6)$$

where $\delta_{\mathbb{X}^i(k)}$ denotes the Dirac mass at particle $\mathbb{X}^i(k)$. We can then approximate the expectation of any integrable function, g , by

$$E[g(\mathbb{X}(k), k)] \approx \int g(\mathbb{X}(k), k) \hat{p}(d\mathbb{X}(k)|\mathbb{Y}(k)) \quad (7)$$

$$= \frac{1}{N} \sum_{i=1}^N g(\mathbb{X}^i(k), k). \quad (8)$$

It can be shown that this estimator is unbiased and (under weak assumptions) converges to the true expectation as the number of particles N tends to infinity [20].

Particle filters suffer from what is known as curse of dimensionality [21] which makes their use in high dimensional

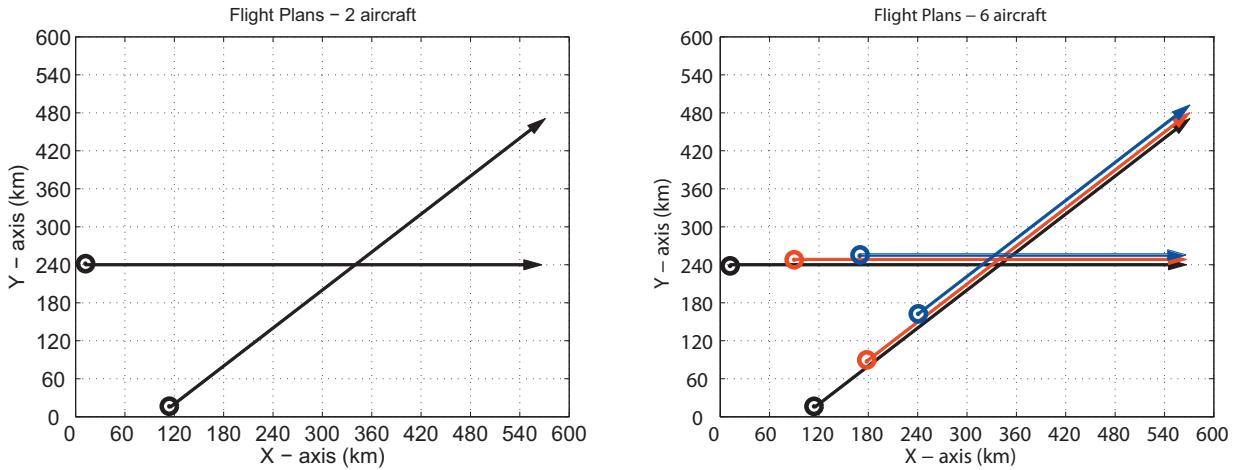


Fig. 4. Flight plans for two (left) and 6 aircraft (right).

problems difficult. Our simulations suggest that most particle filtering algorithms cannot handle efficiently both the high dimensionality of the wind state and the nonlinearities in multi-aircraft dynamics [22].

B. Sequential Conditional Particle Filter

To solve the problems linked to particle filtering we have developed a novel algorithm (which we call Sequential Conditional Particle Filter (SCPF), first introduced in [17]) that can deal with both the nonlinear and the high dimensional nature of the problem. To achieve this we exploit the special structure of the problem, namely the fact that wind forecast error dynamics are linear and that, conditional on the wind, the dynamics of the different aircraft are independent. The SCPF shares some of the insights of the Marginalized Particle Filter (MPF) [23], in the sense, that both treat the linear and the non-linear part of the state separately. The two main novelties compared with the MPF is the sequential incorporation of information from different aircraft and the substitution of particles carrying uncertainty realizations by particles carrying conditional distributions.

The algorithm exploits the fact that the aircraft states evolve according to non-linear dynamics, while the wind states evolve according to linear dynamics. Moreover, the evolution of the wind states is independent of the evolution of the aircraft states and the evolution of the states of different aircraft are only coupled to each other only through the wind states. The first two observations imply that the wind states should be easier to estimate, since under Gaussianity assumptions, storing and manipulating them only requires keeping track of their mean and covariance matrix. Moreover, given a probabilistic estimation of the wind at some points in the wind-field, we can explicitly derive the conditional distribution of the wind at all other points. This distribution will also be Gaussian, hence easy to store and manipulate. This way every aircraft acts as an indirect local sensor of the wind. The first novelty of the proposed algorithm is that, instead

of using realizations for the wind states in our particles (as in conventional particle filtering) we store and manipulate the entire conditional probability distribution. The latter two observations imply that, conditional on the wind states, the states of different aircraft are independent of each other. This is exploited by the second novelty of our algorithm, which is the sequential incorporation of the information from different aircraft. Every radar measurement contains information about the positions, of all aircraft in a region of the airspace, but new measurements are processed one aircraft at the time. The complete algorithm is reported in [17].

IV. SIMULATION RESULTS

We have devised a series of simulations to demonstrate the performance of the new algorithm. Two aircraft approach each other with an angle of 45° . Nominally, without any wind forecast error, the two aircraft exhibit minimum separation (5nmi) after 25 minutes of flight. To demonstrate the algorithm can also exploit information from additional aircraft that happen to be present in the airspace we have created an additional scenario with 6 aircraft in total. The flight plans for the two cases can be seen in Figure 4.

Aircraft fly level at 10000m altitude with a nominal airspeed of 419 knots, and there are no turns included in the flight plans. The parameters of the dynamical models for all aircraft represent a Boeing 737-700. Flights have a duration of approximately 30 minutes and radar measurements arrive every 30 seconds. We simulate these flights under 1000 different wind forecast error realizations. Figure 5 demonstrates the significance of the forecast error. The different scenarios exhibit, on average, their minimum separation after 25 minutes (1500s) as in the nominal case, but the range is now from 1440 to 1570s. For the same flight plan, there exist wind forecast errors for which the separation drops to 0.03 nmi and others for which it reaches 14 nmi. In total, out of the 1000 scenarios, 509 will result in conflict (conflict is 5nmi).

In order to benchmark the efficiency of the algorithm we

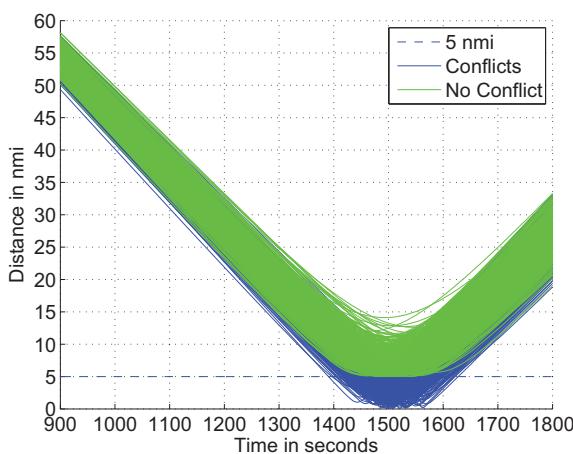


Fig. 5. Time evolution of the separation between the two aircraft for different scenarios of wind forecast errors. Blue lines represent conflicts and green lines represent scenarios where a safe separation distance was kept throughout the flight.

compare with two benchmarks. The first one (called “agnostic”) evaluates the conflict detection rates with only wind forecasts available, without any filtering performed. This is the general case in current practice. The second (called ‘perfect’) evaluates the conflict rates given perfect information about the wind-field all over the airspace at the current point in time. The TP uncertainty in this case is only due to the time evolution of the wind-error, which is unpredictable given the current state. This is clearly an unrealistic, perfect filtering situation and constitutes an optimal, best case performance bound for the proposed algorithm.

The algorithm is run using 1000 particles for the 2 and 6 aircraft case for each of the 1000 wind forecast error scenarios. Having used the algorithm to filter 5, 10 and 15 minutes of data (so 20, 15 and 10 minutes before estimated time of minimum separation) we extrapolate the state estimate of all particles into the future to get an estimate of the future trajectory for each aircraft. We calculate, for each particle the distance of the two aircraft for every second and evaluate the minimum separation throughout the flight. Since each particle will exhibit a different minimum separation, there will be particles for which a conflict has occurred and particles with no conflict for the same wind forecast error scenario. We use the ratio of particles with conflict over the total number of particles as our estimate for the probability of conflict for this scenario.

The results are presented as distributions, using histograms. For each scenario, the algorithm estimates a probability of conflict. We display the results for the scenarios where a conflict actually happened in Figure 6. A distribution skewed to the right implies that most of the conflicts (out of the 509) were identified with high probability. The perfect case would be all 509 cases in the 95 – 100% probability bin. For the scenarios that are placed in each bin the average minimum separation is computed. Minimum separation is defined as the

minimum distance throughout the flight, for the real scenarios.

The agnostic case predictions, Figure 6(a), for 20 and 15 minutes before estimated conflict provide a quite flat distribution. This implies that most of the conflicts are not identified with high probability. The improvement over the agnostic case even only after 5 minutes of flight (20 minutes before conflict) is quite strong for 2 aircraft and becomes even greater when we employ 6 aircraft. This continues for 10 and 15 minutes of filtering and shows the contribution of more aircraft in the airspace. However, after 15 minutes of filtering, adding more aircraft does not significantly increase the conflict detection rate. It is important to note that the 6 aircraft case is not far from the upper margin of performance indicated by the perfect information case, Figure 6(d). Note that even 10 minutes before the conflict there still exist conflicts that are not well identified. SCPF for some cases provides an estimate of the probability of conflict as low as 10-20%. This is due to the conflict being marginal, between 4.5 and 5nm, Figure 6. Accepting conflicts that breach 5.5 or 6nm, would increase margins, with an increase of false alarms of course, but in some cases, this might be a suitable trade-off.

Finally, Figures 7, 8 (including all wind realizations) show how the algorithm improves the estimation of the minimum separation and the time at which it occurs. Standard deviation for minimum separation error is 1.03 nmi and 11s for time error in the agnostic case, while this improves to 0.36 nmi and 2.7s for the 6 aircraft SCPF case. The ideal bound is 0.18 nmi and 1.9s for the perfect information case.

By choosing a probability threshold after which a scenario is considered a conflict we can also estimate the false alarm and successful alert probabilities. The following table shows the result 10min before conflict for a threshold of 90%.

90% Threshold (Agnostic - 6-SCPF - Perfect)		
	Conflict	No Conflict
Alert	51% - 80% - 94%	51% - 22% - 7%
No Alert	49% - 20% - 6%	49% - 78% - 93%

We observe quite a significant increase in the successful alarms and respectively a decrease in false alarms when the SCPF with 6 aircraft is employed, over the agnostic case.

V. CONCLUSION

A method for improving conflict detection (CD) was presented. The performance of the algorithm was tested in flight plans including 2 and 6 aircraft. CD was improved considerably using SCPF compared with the agnostic case where no inference about the forecast error was made. The proposed method manages to both increase the successful alerts and reduce false alerts. Finally, simulations show how the error in the estimates of minimum separation and time to minimum separation are reduced.

ACKNOWLEDGMENT

The authors gratefully acknowledge the contribution of the European Commission through project iFLY, FP6-TREN-037180. The authors would like to thank K. Koutroumpas for helpful discussions and comments.

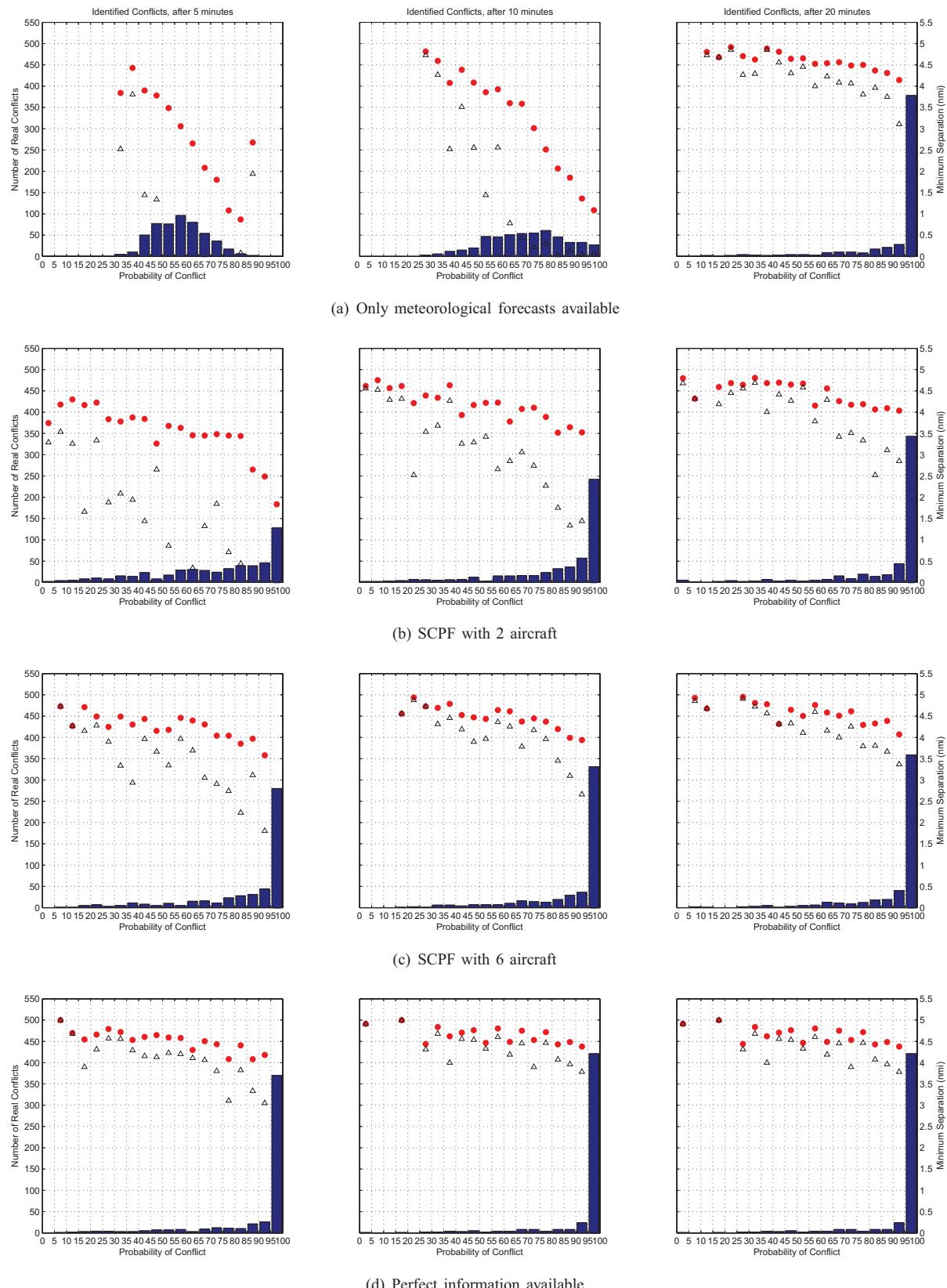


Fig. 6. Evolution of the conflict probability after 5 (left), 10 (middle), 15 (right) minutes of filtering, for scenarios where a conflict occurred. Red dots show the average minimum separation distance for each percentage bin and black triangles the minimum among them - 500 signifies 5nmi for this metric.

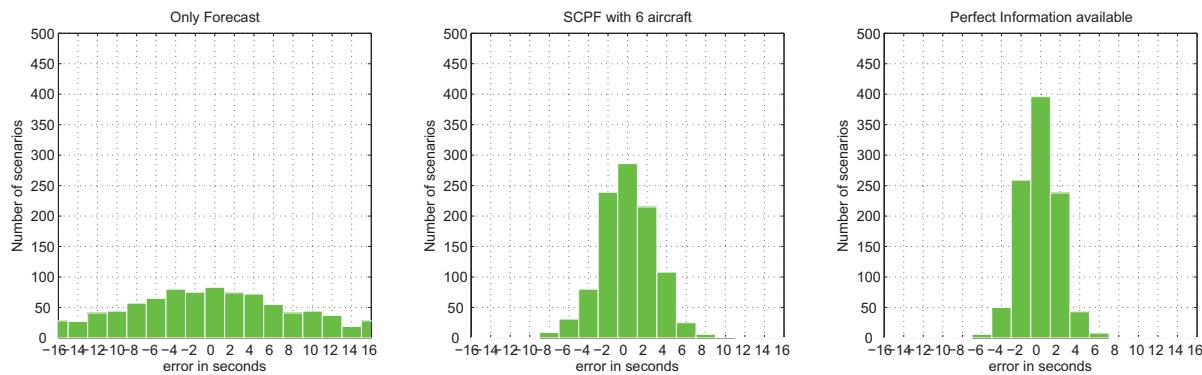


Fig. 7. Distribution of the time error to minimum separation after 15 minutes of filtering for different algorithms

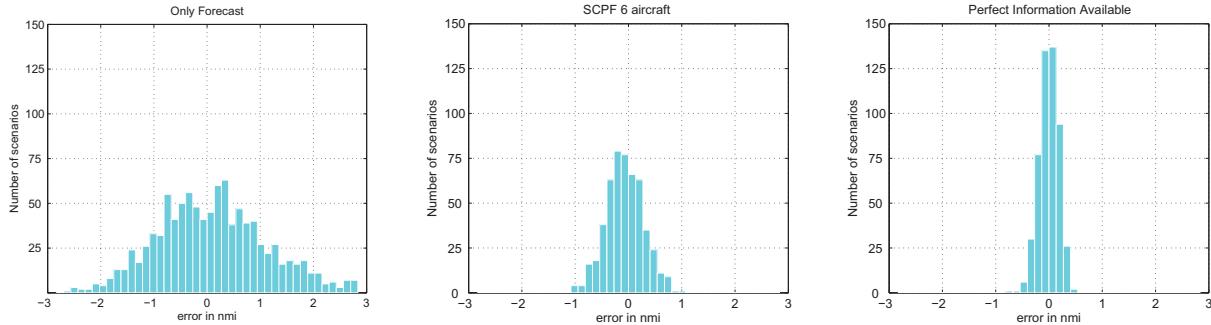


Fig. 8. Distribution of the distance error for minimum separation after 15 minutes of filtering for different algorithms

REFERENCES

- [1] M. Nolan, *Fundamentals of Air Traffic Control*, 3rd ed. Belmont, CA: Wadsworth: Wadsworth Publishing Company, 1998.
- [2] Department of Transport, "UK Air passenger demand and CO₂ forecasts 2009," UK, Tech. Rep., January 2009, <http://www.dft.gov.uk/pgr/aviation/atf/co2forecasts09/>.
- [3] SESAR Joint Undertaking, "SESAR (Single European Sky ATM Research)," 2007, <http://www.sesarju.eu/>.
- [4] NEXTGEN, "Concept of Operations for the Next Generation Air Transport System," 2007, http://www.jpdo.gov/library/NextGen_v2.0.pdf.
- [5] G. Cuevas, I. Echegoyen, J. García, P. Cásek, C. Keinrath, F. Bussink, and A. Luuk, "Autonomous Aircraft Advanced (A3) ConOps," iFly Project, Tech. Rep. Deliverable D1.3, 2008.
- [6] J. Kuchar and L. Yang, "A review of conflict detection and resolution methods," *IEEE Transactions on Intelligent Transportation Systems*, vol. 1, no. 4, pp. 179–189, 2000.
- [7] I. Lymeropoulos and J. Lygeros, "Adaptive Aircraft Trajectory Prediction using Particle Filters," in *AIAA Guidance, Navigation and Control Conference and Exhibit*, Honolulu, Hawaii, August 2008.
- [8] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear estimation*, ser. Information and System Sciences. Prentice Hall, NJ, 2000.
- [9] A. Doucet, J. F. G. D. Freitas, and N. J. Gordon, *Sequential Monte Carlo Methods in Practice*, ser. Statistics for Engineering and Information Science. New York: Springer Verlag, 2001.
- [10] W. Glover and J. Lygeros, "A stochastic hybrid model for air traffic control simulation," in *Hybrid Systems: Computation and Control*, ser. LNCS, R. Alur and G. Pappas, Eds. Springer Verlag, 2004, no. 2993, pp. 372–386.
- [11] I. Lymeropoulos, A. Lecchini, W. Glover, J. M. Maciejowski, and J. Lygeros, "A stochastic hybrid model for air traffic management processes," University of Cambridge, Tech. Rep. CUED/F-INFENG/TR.572, February 2007.
- [12] Eurocontrol Experimental Centre, "User Manual for the Base of Aircraft Data (BADA) revision 3.3," 2002, <http://www.eurocontrol.fr/projects/bada/>.
- [13] S. G. Benjamin, B. E. Schwartz, and R. E. Cole, "Accuracy of ACARS wind and temperature observations determined by collocation," *Weather and Forecasting*, vol. 14, pp. 1032–1038, 1999.
- [14] M. R. C. Jackson, Y. J. Zhao, and R. A. Slattery, "Sensitivity of Trajectory Prediction in Air Traffic Management," *AIAA Journal of Guidance, Control, and Dynamics*, vol. 22, no. 2, pp. 219–228, 1999.
- [15] Earth System Research Laboratory, "Rapid update cycle (RUC)," 2007, <http://ruc.noaa.gov/>.
- [16] R. E. Cole, C. Richard, S. Kim, and D. Bailey, "An assessment of the 60 km rapid update cycle (RUC) with near real-time aircraft reports," MIT Lincoln Laboratory, Tech. Rep. NASA/A-1, July 15, 1998.
- [17] I. Lymeropoulos and J. Lygeros, "Improved Multi-Aircraft Ground Trajectory Prediction for Air Traffic Control," *AIAA, Journal of Guidance Control and Dynamics*, vol. 33, no. 2, pp. 347–362, 2010.
- [18] A. H. Jazwinski, *Stochastic processes and filtering theory*, ser. Mathematics in Science and Engineering. New York: Academic Press, 1970, vol. 64.
- [19] M. S. Arulampalam, S. Maskell, N. J. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.
- [20] D. Crisan and A. Doucet, "A survey of convergence results on particle filtering methods for practitioners," *IEEE Transactions on Signal Processing*, vol. 50, no. 3, pp. 736–746, 2002.
- [21] F. Daum and J. Huang, "Curse of dimensionality and particle filters," in *Aerospace Conference, 2003. Proceedings. 2003 IEEE*, vol. 4, Montreal, Quebec, Canada, September 2003, pp. 1979–1993.
- [22] I. Lymeropoulos and J. Lygeros, "Sequential monte carlo methods for multi-aircraft trajectory prediction in air traffic management," *International Journal of Adaptive Control and Signal Processing*, 2010, to appear.
- [23] T. B. Schön, F. Gustafsson, and P. J. Nordlund, "Marginalized particle filters for mixed linear/nonlinear state-space models," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2279–2289, 2005.

A Stochastic Model for Air Traffic Control Radio Channel Utilization

Vlad Popescu, Henri Augris, Karen Feigh

School of Aerospace Engineering

Georgia Institute of Technology

Atlanta, GA 30332-0150, USA

Email: {vlad.popescu, haugris3, karen.feigh} @gatech.edu

Abstract—This paper offers a stochastic model for radio channel utilization in air traffic control. A log-normal probability distribution for the interaction frequency ('interarrival' times between successive radio interactions), as well as a multivariate joint probability distribution for speech/silence timed sequences have been constructed empirically from over 1,300 hours of recorded radio communication. While the density of communication is nonlinearly reflective of the density and type of traffic, its sequencing and other emerging patterns are also consequences of the controller cognitive and task processes. Such a model for radio channel utilization is a step toward adjusting automated conflict avoidance and decision support tools to effectively match controller attention patterns and task loading. By tuning the output of automated aids to more closely coincide with flight instructions in a manner easily comprehended and perceived by human controllers, this will help reduce workload, improve situation awareness, and address safety concerns by insuring effective human monitoring and trust of future automated maneuvers. Applications may be found in the use of automated advisory assistants suggesting flight instructions in humanly acceptable, albeit mathematically suboptimal, temporal patterns.

Keywords: *air traffic radio communication; frequency utilization; stochastic model; automated conflict resolution advisor; decision support tool*

I. INTRODUCTION

By 2025, U.S. air traffic is predicted to increase two- or threefold, reaching gridlock under the current air traffic practices, and similar forecasts of congestion threaten the European skies. It is widely believed that to accommodate this growth a paradigm shift is necessary; one that will alter the current role of air traffic control from management of traffic by default to management by exception - where the majority of the work required to deconflict traffic will be performed by an automated system working in conjunction with a human controller. In fact, most possible future solutions that have been presented involve automated conflict resolution algorithms, with various levels of implementation and control: from systems where the controller retains responsibility and decision (Wangermann and Stengel, [1]) to robust fully autonomous free flight systems (Masci et al., [2]). It is reasonable to expect that as we move away from current practices, the first step in a more automated ATC paradigm will include automation which provides deconfliction suggestions and acts as a decision aid for human controllers. In that environment the roles of the pilots and air traffic controllers remain as they are today and

the fixed route structure is largely intact. This paper is aimed at such a context and works toward improving these automated assistants from a cognitive engineering perspective.

While it is clear that the introduction of automation will be required, studies such as those conducted by Parasuraman et al. have revealed that hitherto, increase in automation has not been matched by comparable improvements in performance, and the same is to be expected in the future unless care is taken [3]. Human operators often underutilize or conversely overly rely on automation, and Dzindolet et al. have shown that trust is an important factor in understanding automation reliance decisions [4]. If the desired shift in air traffic management is to take place, it appears critical that the automation put into place be transparent, reliable, and synergistic with its human operator's work practices.

In the present as well as in the future context we are describing, changes to flight plans will be issued in the form of clearances by the controller and confirmed by the pilot. For this type of automation, which provides conflict resolution advice to the controller for approval, several questions regarding automation functionality arise:

- How often should advice be provided?
- How much advice should be provided at any time?

We believe that a step toward answering these questions requires understanding the cadence of controller-pilot communication. The temporal spacing of flight instructions is reflected in communication patterns, and hence the study of current communication may provide a valuable means of extrapolating controller work rhythms. By taking into account the customary manner that controllers have of interacting with air traffic management, automation will be less disruptive.

Specifically, an understanding of the temporal spacings in pilot-controller communication is needed in order to adapt the output of such automation to provide maneuver suggestions in a sequence corresponding to the human controller's attention and perception patterns. The cognitive term 'workload' has a broad span, encompassing a measure of the physical number of tasks to be completed (taskload) along with other factors such as skill and stress. This is the definition we are referring to in this paper. For a controller assisted by an automatic decision advisor, respecting the work pattern will help reduce workload and improve situation awareness.

Controller-pilot radio frequency utilization is a relatively recent topic of interest from the perspective we are discussing here, and has only been tangentially addressed (cf. literature review by Prinzo and Britton [5]). Radio communication has mostly been a focus of research aimed at analyzing complexity, miscommunication and sources of errors. Burki-Cohen has examined the incidence of the complexity of controller instructions on communication problems and conceptual errors [6]; Prinzo, Hendrix and Hendrix have shown that message complexity is correlated to errors of omission, while message length affected both the production of errors of omission and readback errors [7]; Prinzo and Morrow have elsewhere attempted to improve understanding by studying the influence of message format and message length [8]; Prinzo has also categorized the communication elements according to their semantic taxonomy and the corresponding distribution of ambiguities or misunderstandings [9].

Rakas and Yin have found that pilot initiated communications at arrival into or departure from a sector are more likely to cause errors than those initiated by controllers, that departures rather than arrivals cause more extensive and significant miscommunications, and suggest that there is less miscommunication in situations when a controller is handling several aircraft simultaneously [10]; Cardosi has conducted similar studies for the purpose of examining communication practices and the relation between complexity and error in the enroute [11] and terminal (tower) [12] environments; Howard's analysis of problematic communication showed that pilot speech contained more errors than controller speech, that higher amounts of information being transmitted in one sentence led to increased problematic communication in subsequent interactions, as did linguistic violations of protocol [13].

Straussberger has proposed a model of monotony that considers repetitiveness and uneventfulness, the individual boredom proneness and states at the beginning of the work shift as well as organizational factors to assess monotony with the help of physiological, subjective, and behavioral indicators, in distinction with other states such as fatigue and satiation [14].

Some authors have also explored workload and its relation to voice communication patterns and channel occupancy. Porterfield's work has shown high correlation between the controllers' communication duration and the controllers' subjective workload, and validated communication duration as a measure of workload [15]; Prinzo et al. indicate that hesitation pauses in speech occur in low workload conditions, when the controller has time to think and responds in a cognitive way, as opposed to an automatic way encountered when a large number of aircraft are active on the frequency [16]; Bolic et al. have conceived a cognitive utilization metric, meant to capture the ratio of time in which a controller thinks about certain aircraft, and have found this to be correlated with the utilization of the radio channel. Nonetheless in cases when the former exceeds the latter, controllers' and pilots' capacity to

conduct voice communications is substantially reduced [17]. Unfortunately, Manning et al. have shown that the addition of communication to a model of workload does not improve or render more precise its estimates deduced from conventionally measured air traffic control taskload data [18].

By correlating communication data with vector deviations from the assigned flight paths, Yenson and Rakas have developed a controller workload model demonstrating increased efficiency with the use of a mixed voice and datalink system environment that reduces voice channel usage and by extension controller workload [19].

What is lacking in this literature is a quantitative description of pilot-controller interaction that would allow designers to integrate information on communication patterns into automated tools. This paper seeks to develop a stochastic model of air traffic control and aircraft pilot communication patterns from empirical data. The ambition of such a model is ultimately that of enhancing air traffic control automation and decision aids by matching the output with pre-existing human controller communication rhythms and attention patterns. It is therefore important to note that our focus is not the direct improvement of the current work practices in air traffic control, nor the theoretical characterization of channel capacity from an information theory standpoint, but rather the improvement of future automation such that it will feel more natural to the human it is designed to assist. We suggest that the integration of a type of automation cognizant of human idiosyncrasies would benefit its acceptance, its intelligibility, and fundamentally the reliance on it. We do however acknowledge the limited range that comes from measuring the controller's perceptions with no reference to the actual traffic flow patterns, along with the potential bias that might be conveyed. Extending the method showcased here to new traffic situations will require that it be put into context through the characterization of flow patterns (see Section V).

II. ANALYSIS

The construction of an inter air traffic control - pilot communication model began with the collection of communication data. An analysis in three stages was conducted to identify the statistical distribution of the communication parameters.

A total of 1,326 hours of radio communication related to Atlanta's Hartsfield-Jackson airport and northeast arrivals has been analyzed. Specifically, there are 344 hours for the TRACON frequency at 127.25 MHz (North final approach into ATL), 381 hours for the enroute center frequency at 121.35 MHz, covering sector 49 ('Logen' - represented in Figure 1). Logen is Atlanta's northeast low altitude sector handling all arrival aircraft from the Northeast between 11,000 feet and 23,000 feet), 344 hours for the ATL tower frequency at 119.1 MHz and 257 hours for the ATL ground frequency at 125.32 MHz.

For comparison purposes, additional frequencies from New York City's JFK Airport have also been analyzed, over a total of 184 hours. There are : 76 hours for JFK tower broadcasting

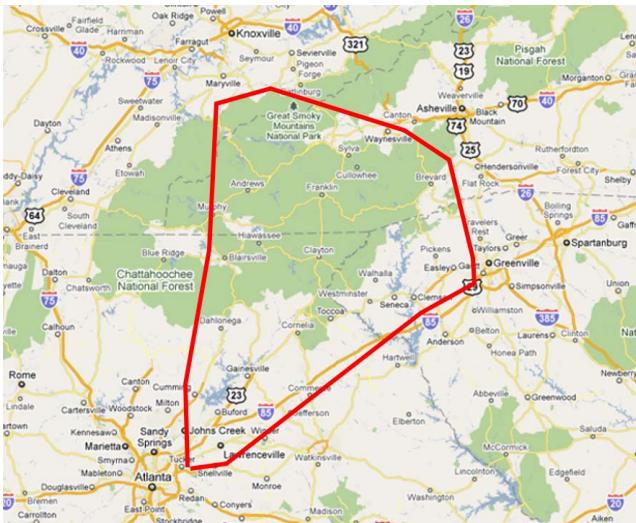


Fig. 1. Atlanta low-altitude sector 49 (Logen) Map data ©2009 Google

on 119.1, 123.9, 125.25 MHz, 37 hours for JFK approach (127.4 MHz), 30 hours for JFK final approach (132.4 MHz) and 41 hours for JFK north ground (121.9 MHz).

A Matlab analysis was conducted in three stages; the first stage consisted of detecting sound defined as an average signal intensity surpassing a graphically determined threshold that separated white noise static from coherent speech, over a 0.15 second window. The second stage of the analysis was meant to classify and time conversations and silence periods (see Figure 2). A silence period was defined as no noise over 1.5 seconds (in other words no sound detected over 10 window periods), and conversations were then assumed to form the complement of the silence periods. Shorter periods of silence were assumed to be natural pauses in speech and merged with neighboring conversations. In the final stage, periods of conversation and silence were assembled into histograms to approximate probability distributions for the signal, either with Matlab or with the ExpertFit [20] distribution-fitting software.

Our choice of silence as the main characteristic was due to its more objective definition. While noise and conversation come in very different and irregular patterns, it was assumed that inter-conversation silence would consist of a continuous lack of coherent signal. The statistical analysis of the channels shows (Table I) varying results for the talk and quiet times balance.

III. STOCHASTIC MODEL : MESSAGE INTERARRIVAL TIME

The variable we believe would be most useful for integration to decision-aiding output is the ‘interarrival time’ (cf. definition in Figure 2 and plots in Figure 3), defined as the duration between the beginning of one conversation and the beginning of the next (in other words the duration of consecutive ‘talk’ and ‘quiet’ times). This interarrival time variable allows a temporal spacing for suggested commands to be established in a way that is coherent with current controller work practices. We recognize that two types of information flow exist, routine

TABLE I
RADIO DATA VOLUME

Frequency	Conversations	Talk time	Silence time	Talk ratio
ATL Center	31847	132 h	249 h	35%
ATL Approach	35185	182 h	162 h	53%
ATL Tower	37993	135 h	209 h	39%
ATL Ground	24963	122 h	135 h	47%
JFK Approach	3881	13	24	35%
JFK Final approach	2360	15	15	49%
JFK Tower	7805	24	52	32%
JFK Ground	4819	13	28	33%

information that is linearly related to traffic flow and conflict resolution information that is exponentially related to it. By this approach, only the aggregate can be measured.

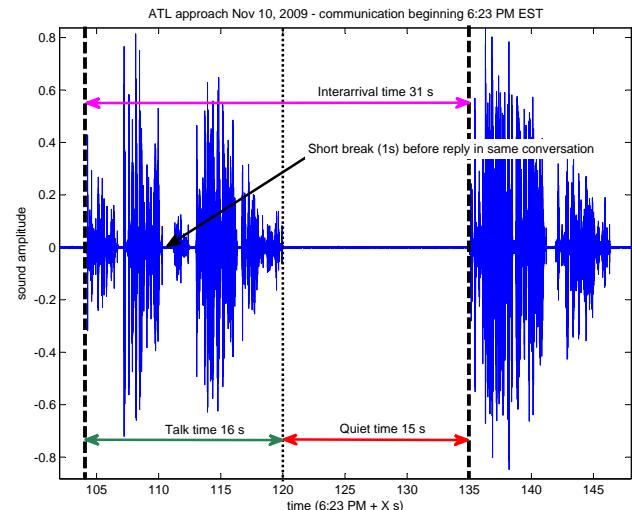


Fig. 2. Interarrival time definition

A least square regression was used to classify the data, not providing an exact analytical match for the probability density function. The analysis of communication interarrival time variable is most closely associated with a log-normal probability density distribution (cf. Figures 4 and 5), with a mean squared error below 0.001. The log-normal distribution may also be verified by plotting the logarithm of the variable (cf. Figure 6) which must obey a normal distribution. This result was not expected, as it shows that the communication interarrival is only partially correlated with the aircraft sequencing in entering the sector. Under the assumption that these events occur continuously and independently of one another, aircraft entry into a sector has been modeled in simulations by a Poisson process in off-peak hours (see Salaün et al. [21]),

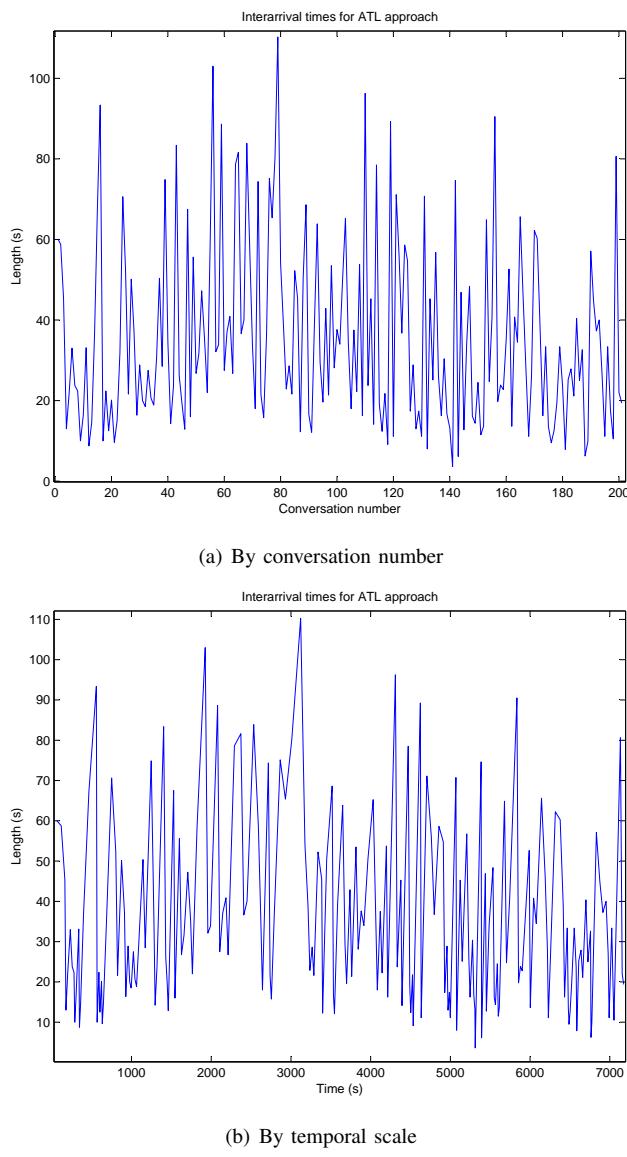


Fig. 3. ATL approach interarrival length

and consequently the corresponding aircraft interarrival times obey an exponential distribution, or a normal distribution in peak hours when the aircraft arrival queue is saturated. As radio communication is highly dependant on (although not directly proportional to) aircraft flow input and output, a closer match between communication and flight sequencing was expected. Nevertheless, we also acknowledge that the time of day (peak hours saturation), the aircraft types (Heavy, Medium, Light), the technology available would influence the communication patterns. In this basic approach, no fluctuation according to time of day or other traffic flow characteristics has been considered, but we do expect such an analysis to reveal nonhomogeneity (see Section V).

The repeated or case-specific interactions between the controller and some of the aircraft may at least partially account for this distinction. Nonetheless, it is also true that the math-

ematical fit tests meant to precisely identify a probability distribution for the communication interarrival time remain inconclusive: the χ^2 test only yields a p-value of less than 0.01. We believe the unrefined manner of analysis to play some part in this: our definition of the sound detection threshold and of the temporal analysis windows (chosen to be 0.15 second for sound identification, 1.5 seconds for silence periods) are inferences and educated assumptions. Only limited empirical verification by audition of the corresponding radio transmissions has been performed at this stage. Insofar, the trend regarding short conversations is confirmed, however it appears that the numerical analysis tends to overestimate longer conversations (over 15 seconds), possibly by assembling several short conversations with distinct aircraft that take place in rapid succession. Refining the numerical method is thus likely to produce results that may come closer to a pure log-normal distribution.

Nonetheless, we find that these results offer a solid general model for the communication on a large temporal scale. As might be expected, the parameters are dependant on the monitored frequency and the corresponding activity. Comparison with the JFK recordings also shows that the convergence of the measured activity toward this distribution is slow, and while a certain uniformity is observed over periods on the order of ten days, that is not the case at the scale of one day. The statistical parameters for the monitored signals are shown in Table II.

TABLE II
STATISTICAL PARAMETERS OF INTERARRIVAL

Frequency	Median	Mean	Coefficient of variation	Skewness
ATL Center	22	43	2.4	12.7
ATL Approach	26	35	2	25.5
ATL Tower	20	33	1.8	12.4
ATL Ground	22	37	1.4	8
JFK Approach	17	35	4.3	25
JFK Final approach	19	46	6	15.3
JFK Tower	20	36	2.2	13.9
JFK Ground	15	31	1.9	10.6

IV. JOINT PROBABILITY DISTRIBUTION : TALK AND QUIET TIMES

We believe that the question of conditional probability, i.e. that of the interdependance of communication on silence periods, should be investigated. The conclusion that communication patterns follow a log-normal distribution relies on large-scale, historical trends that are inherently imprecise regarding any specific behavior or time period. Either because of extenuating circumstances, changing traffic, judgment calls, personal preferences or experience, a controller would not

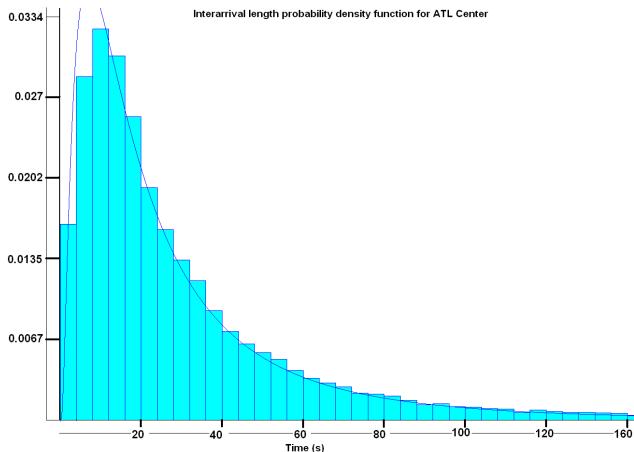
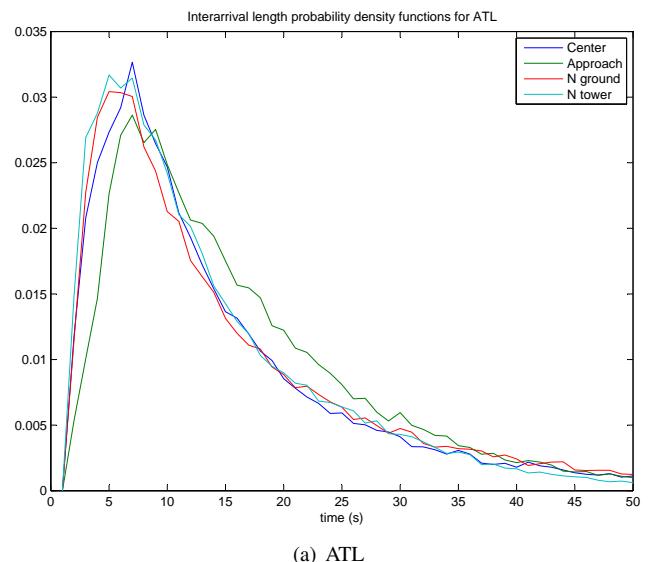
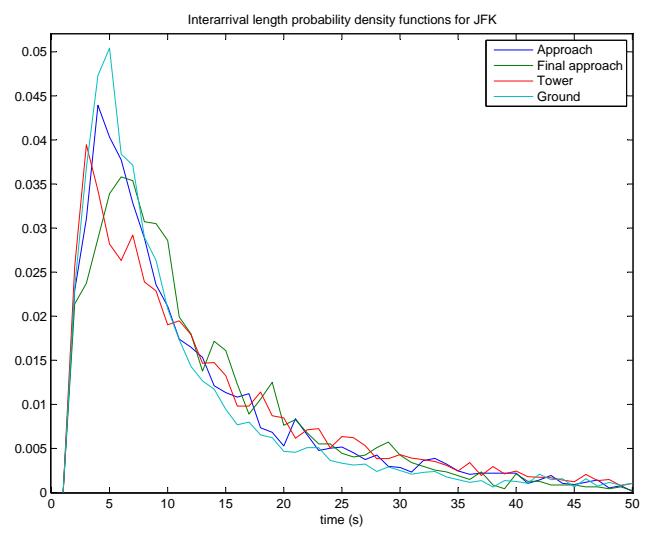


Fig. 4. ATL center interarrival log-normal probability fit



(a) ATL



(b) JFK

Fig. 5. Probability density functions

always apply the communication pattern exactly as predicted and suggested by the automation. It is expected that the work and communication patterns of a controller assisted by decision-aid automation of the kind we are suggesting might stray significantly in the short term from the predicted behavior, i.e. a log-normal distribution, although will respect the long term homogeneity that has been found in this data.

Therefore, in order to account for this variability in communication, a method for dynamic readjustment may prove useful to the automation. That is, the automation may benefit from using the conditional probability of communication and silence in order to better calibrate the timing of its advice. Since a balance must be struck between on the one hand respecting past trends and on the other locally adapting to the changing conditions, adaptability of the automation is the key to reducing controller workload and improving transparency. For that purpose, we have constructed a bivariate joint probability density function for the balance of talk and quiet times, from which any conditional probability may be deduced. The plot in Figure 7 represents $P(\text{talk} = X \wedge \text{silence} = Y)$.

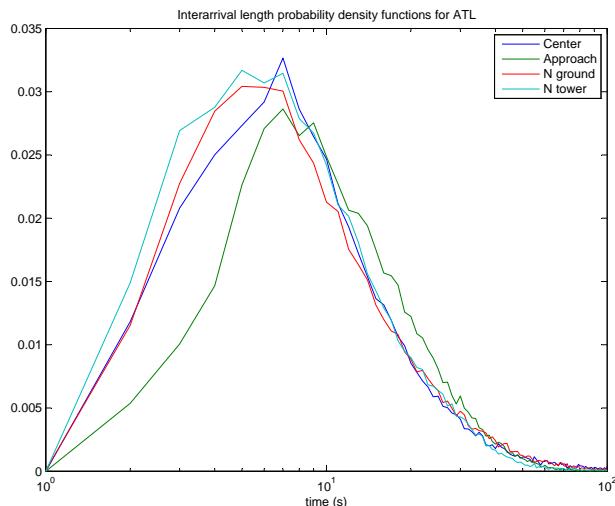
The conversation length and silence length variables appear to be dependant. A Bayesian gap was calculated, defined by $|P(\text{talk} = X \wedge \text{silence} = Y) - P(\text{talk} = X) * P(\text{silence} = Y)|$ (plotted in Figure 9). For independant variables, this must be null. Under the hypothesis that our measurements contain sufficient occurrences to approach the total univariate probability distributions for conversation and silence lengths, we have calculated an approximation of $P(\text{talk} = X) * P(\text{silence} = Y)$ (plotted in Figure 8), and $P(\text{talk} = X \wedge \text{silence} = Y)$ is given by the histogram method (plotted in Figure 7). For the shorter time spans, a significant gap is clearly visible, amounting to 20% or more (notice the gap value of 0.004 for approximated probability values on the order of 0.015). These results are consistent throughout the different analyzed frequencies (cf. Annex); the consecutive talking and silence periods are therefore found to be correlated such that the duration of a conversation will influence the subsequent time

the controller will remain silent, and reciprocally.

V. POSSIBLE APPLICATIONS AND FUTURE WORK

As an immediate application, the results presented here allow a more exact modelling of air traffic control interaction with aircraft for simulations that are used in the validation of conflict resolution algorithms and the evaluation of the workload reduction and/or the efficiency increase brought about by these solutions. Further refined, this stochastic model will allow the tuning of decision aids and automation for use in air traffic control settings.

In the future, following a more precise analysis and more extended audition, we expect to gain sufficient insight to refine our definition of silence / speaking and obtain a more precise outline of the communication patterns along with their purpose. While so far the measurement has been limited to



(a) ATL

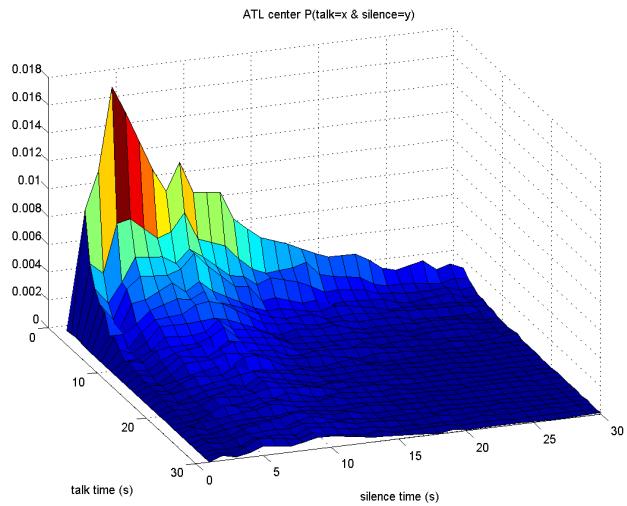
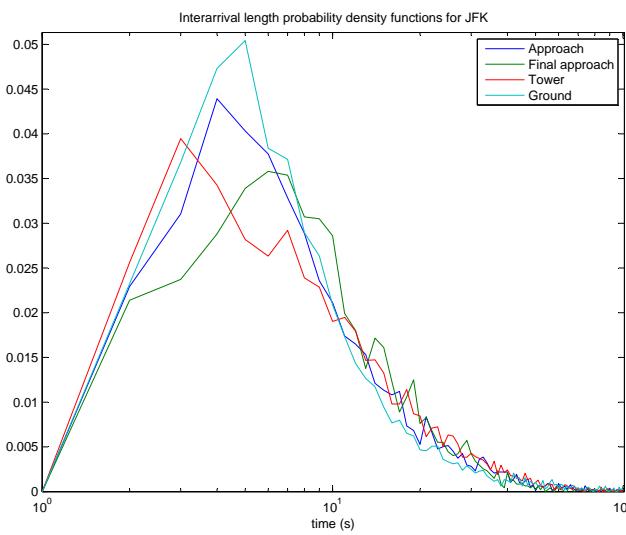


Fig. 7. ATL center measured bivariate probability density



(b) JFK

Fig. 6. Logarithmic scale probability density functions

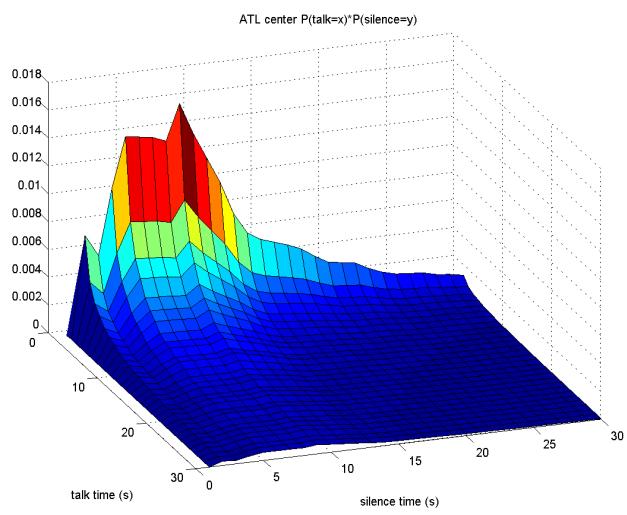


Fig. 8. ATL center approximated bivariate probability density

VI. CONCLUSION

physical channel occupation, our ultimate focus is on semantic utilization. We also expect to clarify the influence played by the frequencies' specific use (e.g. approach, ground,...) and to further study the distribution convergence in relation on the total sample size. Another notable direction we are investigating is the relation between communication patterns and the time of day. Furthermore, a major preoccupation remains the correlation between communication against sector geometry, aircraft flow patterns, flight conflict probabilities, weather, and their respective impact on controller workload. This link is essential for putting communication patterns into context and for generalizability. Work on distinguishing routine communication from conflict resolution interaction is currently underway.

This paper has presented a stochastic model for radio channel utilization in air traffic control. From the analysis of over 1,300 hours of radio communication in Atlanta and New York - JFK, a log-normal probability distribution has been identified for the interarrival times between successive radio conversations, and a multivariate joint probability distribution for speech/silence timed sequences has also been constructed. The implementation of such a model for radio communication in future automation is a step toward adjusting automated conflict avoidance and decision support tools to effectively match controller attention patterns and task loading. Under a new paradigm of automation-assisted or automation-controlled air traffic management, we believe this to be a direction toward reduced workload, improved situation awareness, and the respect of safety concerns.

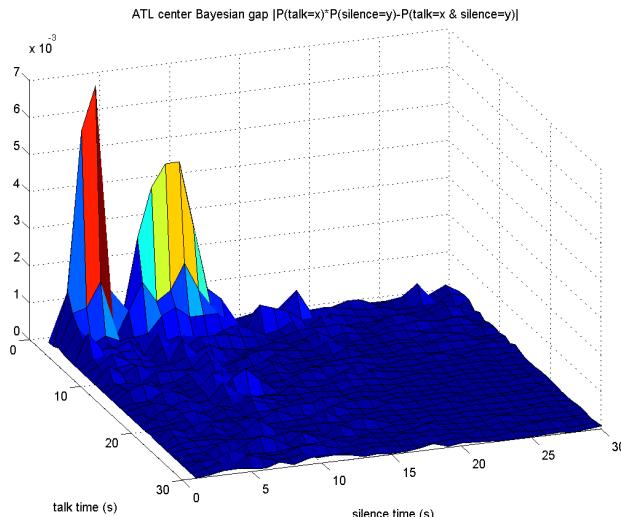
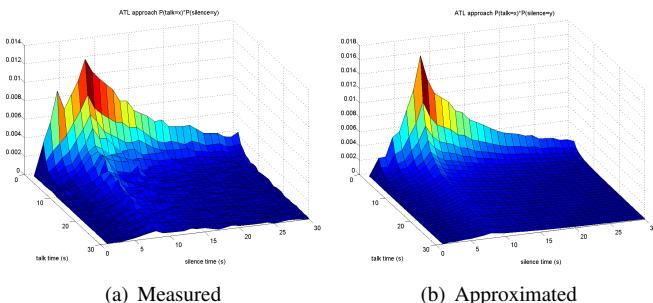


Fig. 9. ATL center Bayesian gap

REFERENCES

- [1] J. P. Wangermann and R. F. Stengel, "Principled negotiation between intelligent agents: a model for air traffic management," *Artificial Intelligence in Engineering*, vol. 12, pp. 177–187, 1998.
- [2] P. Masci, H. Moniz, and A. Tedeschi, "Services for fault-tolerant conflict resolution in air traffic management," in *Proceedings of the 2008 RISE/EFTS Joint International Workshop on Software Engineering for Resilient Systems*, 2008, pp. 121–125.
- [3] R. Parasuraman and V. Riley, "Humans and automation: use, misuse, disuse, abuse." *Human Factors*, vol. 39, pp. 230–253, 1997.
- [4] M. Dzindolet, S. Peterson, R. Pomranky, L. Pierce, and H. Beck, "The role of trust in automation reliance," *International Journal of Human-Computer Studies*, vol. 58, no. 6, pp. 697–718, 2003.
- [5] O. Prinzo and T. Britton, "Atc/pilot voice communications - a survey of the literature," Tech. Rep., 1993.
- [6] J. Buerki-Cohen, "An analysis of tower (ground) controller-pilot voice communications," FAA, Tech. Rep., 1995.
- [7] O. Prinzo, A. Hendrix, and R. Hendrix, "The outcome of atc message complexity on pilot readback performance," Tech. Rep., 2006.
- [8] O. Prinzo and D. Morrow, "Improving pilot/air traffic control voice communication in general aviation," *The international journal of aviation psychology*, vol. 12, no. 4, pp. 341–357, 2002.
- [9] O. Prinzo, "An analysis of approach control/pilot voice communications," FAA, Tech. Rep., 1996.
- [10] J. Rakas and H. Yin, "Analysis and modeling of controller-pilot miscommunication messages," in *AIAA 5th Aviation, Technology, Integration, and Operations Conference (ATIO) 26 - 28 September 2005, Arlington, Virginia*, 2005.
- [11] K. Cardosi, "An analysis of en route controller-pilot voice communications," FAA, Tech. Rep., 1993.
- [12] ———, "An analysis of tower (local) controller-pilot voice communications," FAA, Tech. Rep., 1994.
- [13] J. Howard, "Tower, am i cleared to land?: Problematic communication in aviation discourse," *Human Communication Research*, vol. 34, no. 3, pp. 370–391, 2008.
- [14] S. Straussberger, "Monotony in air traffic controller - contributing factors and mitigation strategies," EUROCONTROL Human Factors Laboratory, Tech. Rep., 2006.
- [15] D. Porterfield, "Evaluating controller communication time as a measure of workload," *International Journal of Aviation Psychology*, vol. 7, no. 2, pp. 171–182, 1997.
- [16] O. Prinzo, P. Lieberman, and E. Pickett, "An acoustic analysis of atc communication," Tech. Rep., 1998.
- [17] T. Bolic, J. Rakas, and M. Hansen, "Controller-pilot radio channel utilization and cognitive issues," in *6th USA/Europe Air Traffic Management R&D Seminar, Baltimore*, 2006.
- [18] C. Manning, S. Mills, C. Fox, E. Pfleiderer, and H. Mogilka, "Using air traffic control taskload measures and communication events to predict subjective workload," Tech. Rep., 2002.
- [19] S. Yenson and J. Rakas, "Impacts of a mixed media air traffic control communication environment on aviation efficiency," in *12th Air Transport Research Society (ATRS) World Conference Athens, Greece, July 6-10, 2008*, 2008.
- [20] Averill M Law and Associates. (2010) ExpertFit Distribution-Fitting Software. Tucson, AZ.
- [21] E. Salaun, M. Gariel, A. E. Vela, E. Feron, and J.-P. B. Clarke, "Statistical proximity maps based on data-driven flow modeling," in *AIAA Infotech@Aerospace*, 2010.

ANNEX: BIVARIATE PROBABILITY DISTRIBUTIONS



(a) Measured (b) Approximated

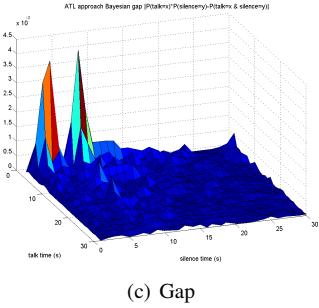
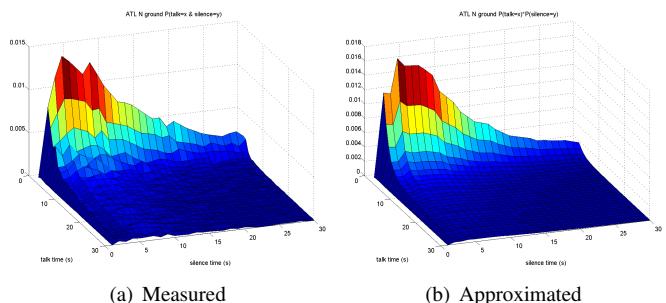


Fig. 10. ATL approach



(a) Measured (b) Approximated

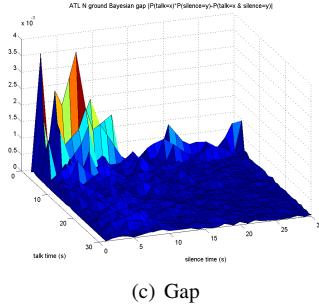
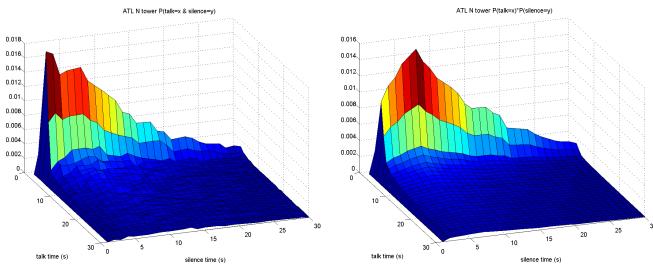


Fig. 12. ATL ground



(a) Measured (b) Approximated

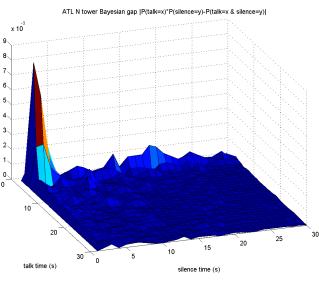


Fig. 11. ATL tower

Impact of US Airline Network Topology on Air Transportation System Performance

Tatsuya Kotegawa, Donald Fry, Ethan Puchaty, Daniel DeLaurentis

School of Aeronautics and Astronautics

Purdue University

West Lafayette, IN 47907 USA

tat@purdue.edu, dfry@purdue.edu, epuchaty@purdue.edu, ddelaure@purdue.edu

Abstract— Much of the current research aimed at reducing the air transportation system's impact on the environment revolves around increasing the aircraft fuel efficiency or improving air traffic management practices. There are, however, many other factors that play a role in determining the system-wide performance of air transportation, such as the airline service route network topology characteristics, aircraft fleet mix and resource allocation. This paper investigates the impact of different service route network topology types on transportation efficiency and robustness metrics developed by the authors.

Keywords-network; efficiency; airlines; environment; system-wide performance

I. INTRODUCTION

Transforming the national and international Air Transportation Systems (ATS) to meet future travel demand has been the focus of many researchers and decision-makers. This challenge has become further complicated by increased noise and emissions restrictions stemming from growing awareness of the aviation industry's impact on the environment and by increased economic pressure due to volatile fuel prices. Improving individual aircraft efficiencies and air traffic management (ATM) practices have been common approaches to satisfy increasing travel demand while reducing environmental impacts. While it is very important to assess and improve the ATS performance at the level of individual aircraft and ATM procedures, there are many other high-level factors beyond these that determine the system-wide performance of air transportation, such as the airline service route network topology, aircraft fleet mix and resource allocation. These factors are extremely large in scope and their complex nature makes analysis as well as subsequent design decisions extremely difficult.

The lack of a universal definition that describes the overall performance of the ATS exacerbates the problem. This is mainly due to the distributed control and heterogeneous structure of the ATS composed of multiple stakeholders (e.g., passengers, airlines, airports, etc.) operating under a unique set of objectives, timescales and domains (e.g., economical, operational, and political) [1]. Since each stakeholder has their own set of objectives, they also have their own perception of what "ATS performance" means. For example, ATS performance for an airline may be based on the economical effectiveness of meeting passenger travel demand. However,

from a passenger point of view, performance may also be based on required travel time or number of connections, which does not necessarily coincide with an ATS architecture designed for economical effectiveness (e.g., hub-and-spoke type service route network). Further, ATS performance defined by the amount of noise or emissions released may contradict with metrics formed for either the passengers or the airlines.

The research reported in this paper describes the preliminary steps taken by the authors in analyzing the trade-offs between performance metrics from different stakeholder standpoints. More specifically, this paper investigates trade-off studies between passenger-centered efficiency metrics and system robustness for different types of airline service route network topologies. The remainder of the paper is organized as follows: After a brief literature review on some of the performance metrics related to the ATS in Section II, network theory is introduced. Section III describes the metrics that were used as a baseline to compare the performance of the various service network topologies, portrayed in Section IV. Section V summarizes the interim results, followed by key implications in Section VI.

II. BACKGROUND

A. Literature Review

In recent years, a significant amount of research towards improving the ATS performance was based on improving individual aircraft performance. Reference [2] is an example of this, employing energy usage and specific energy intensity—largely aircraft-centric measures—as the performance metrics for analyzing the current and historical ATS. Energy usage and specific energy intensity were examined as a function of different types and classes of aircraft. These metrics are primarily influenced by aircraft design decisions, such as propulsion type, passenger load, technological evolution, and the specific mission design requirements. By using energy as part of their metric, the authors implied that improving the fuel efficiency of an aircraft would have a direct impact on the overall air transportation network efficiency. This allowed for the impact of individual aircraft design parameters on the overall efficiency of the air transportation system to be explored. Since passenger load was also taken into account via specific energy intensity, some light was also shed on the effects of fleet operations on

transportation efficiency. However, these metrics do not necessarily provide explicit results for changes in specific fleet operations or network topology. In addition, there are trade-offs between airline equity and passenger equity, as Manley and Sherry [3] demonstrated.

Instead of individual aircraft metrics, [4] used fleet level metrics to examine how changes in fleet operations affect air transportation network performance. Route demand, number of aircraft on route, route distance, passenger load, number of aircraft, and maintenance hours were among various factors used to create an objective function modeling fleet efficiency. These factors were primarily affected by the fleet distribution and allocation of different types and numbers of aircraft to each route. It also presented methods to implement new aircraft technology into the tool to obtain new and ideal fleet distributions, thereby linking aircraft design with fleet efficiency. Again, while changes in the fleet mix were examined and applications to aircraft design were offered, changes in network topology were not formally addressed.

Reference [5] addressed network utilization by examining the cost of establishing routes based on an efficiency metric that examined the trade-off between the wait/fly ratio and route distance ratio. Using airlines as rational agents, the wait/fly ratio and route distance ratio were weighted and the cost and utilization of the ATS were evaluated. While this allowed for the comparison of network properties with performance, the process was not directly related to the aircraft design process, and again network topology was not necessarily an intended design parameter.

Two tools currently under development served as inspiration and background for this paper: The Aviation Environmental Portfolio Management Tool (APMT) [6] and Aviation Integrated Modeling Tool (AIM) [7]. Both of these tools were and are being used in the context of evaluating system benefits, costs, policies, operations, etc. based on a set of inputs specified by the user. The purpose of each tool was to provide the user with options for changing system inputs, parameters, and characteristics in order to achieve an efficiency goal. What set these two tools apart from aircraft-centric research on the topic of efficiency was their use of network architecture and complex layers of objectives as part of the analysis. However, while these tools addressed the fact that small changes in the network architecture could result in large scale differences in ATS performance, no specific measure of ATS efficiency was presented or validated as a proper baseline. This provided the motivational basis and outline for the experimental design in which the network topology was varied in order to achieve an performance target, as described in subsequent sections.

B. Introduction to Network Theory

Network Theory has produced powerful results from multiple domains (e.g., physics, information, social science, biology) in recent years concerning how real-world networks are structured. Some researchers have applied the analysis techniques developed in the network theory community to explore the structure of the ATS. Guimera, et al. analyzed the worldwide air transportation network topology and computed measures which characterized the relative importance of cities

and airports [8]. Further, Bonnefoy and Hansman used the weighted degree distribution for light jet operations to understand the capability of airports to attract the use of very light jets [9]. A significant body of works exists in the related domain of operations research on the design of optimal networks for particular instances and applications (e.g., schedule for an airline). However, these approaches generally do not pursue how the underlying network topology influences the characteristics of the ATS as a whole, the interplay between networks that reside in different domains, nor the role these structures play in future designs. Applying network theory not only as an analysis tool but also for designing the future ATS has been a continuing topic for our work [10, 11]. In particular, this paper examines the trade-offs between efficiency and robustness (can also be thought as risks) of different network topologies for the airline service route network is investigated in Section IV and V, which may be applied towards future ATS designs.

III. TRANSPORTATION EFFICIENCY FORMULATION AND ANALYSIS OF THE HISTORICAL ATS

In systems with multiple stakeholders such as the ATS, objectives between the stakeholders may conflict. As a result, metrics may favor one stakeholder over another in representing how efficient the system is. Further, stakeholders seek to optimize their operation with regard to their own objectives to maximize benefits from the system. For instance, airlines use the hub-and-spoke system to reduce costs. In particular, hubs allow airlines to aggregate passenger origins to more efficiently transport them to their destinations. While this may be ‘efficient’ for the airlines from an economic standpoint, it may be detrimental to the passengers in that they may need to travel extra distance to their destination, or for the regulators that prefer to keep the density of operations and ATM workload low to maintain safety.

As an initial step to investigate the potential trade-offs between efficiency from various stakeholder’s standpoint, passenger travel distance efficiency was created to investigate the impact of airline service route network topology on travel distance for passengers, shown below.

$$\tau = \frac{d_{ij}}{d_{tot}} \quad (1)$$

d_{ij} is the distance between the passenger's origin and destination where d_{tot} is the total distance traveled by the passenger, which includes connections, if any. In this formulation, τ is less than or equal to one, where $\tau = 1$ for direct flights.

Using this formulation, τ was calculated for every itinerary in the DB1B datasets from 1993 to 2007. The DB1B is a 10% sample of all itineraries flown and reports the actual routing of passengers. Calculating τ for each passenger allowed the average efficiency to be computed. Efficiency varied from year to year, as shown in Figure 1, but averaged quite high at 92.7% for the years investigated. To examine some of the potential effect of airline’s hub-and-spoke structure, the

average efficiency was also computed for indirect flights only. As expected, these efficiencies (also shown in Figure 1) were lower, but only slightly, averaging 88.5% for the 15-year period.

According to the data, the average passenger flying on an itinerary with at least one connection traveled 12.5% farther than he/she would have on a direct flight. Given that the average trip length (from origin to destination) in 2007 was 917 km (570 mi), the average traveler flew an extra 119 km (74 mi). Note that this is only an indication of how much farther a traveler was made to travel and does not take into consideration monetary cost or additional time required from connections to the traveler.

As different airlines route passengers differently according their service network, it stands to reason that the τ of one airline may differ from another. To explore this, we selected two types of airlines: traditional hub-and-spoke carriers (Delta and American Airlines) and an airline with more point-to-point operations (Southwest Airlines). Using the same method for calculating τ for all flights, time histories of average τ were created

Delta had slightly lower efficiencies, as shown in Table 2, but followed the same general trend as the overall ATS. Southwest Airlines, with its focus on point-to-point service had nearly equivalent efficiencies (to the overall ATS) over the 1993–2007 timeframe. However, Southwest was also growing its operations during the first half of that period and still operated many of their flights from a few major airports. As their service area grew, more direct flights were added, resulting in slightly higher τ for travelers since 1999, as shown in Figure 2. Considering only itineraries from 2001 to 2007, τ of the average passenger flying on Southwest was 2% higher than the national average (see Table 1).

TABLE 1. PASSENGER TRAVEL DISTANCE EFFICIENCY ACCORDING TO AIRLINE

Airline	Average Passenger Excess Travel Ratio (τ)	
	All Flights	Indirect Flights
All Airlines	92.7%	88.5%
Delta Airlines	90.8%	87.4%
American Airlines	92.9%	88.9%
Southwest Airlines	93.2%	88.6%
Southwest Airlines (2001–2007)	94.8%	89.4%

In addition to Delta and Southwest, flights on American Airlines were also analyzed. This allowed service network topology and carrier operations to be compared with respect to τ . In this case, Delta represented an airline with a hub-and-spoke structure which also participated in a large degree of code-sharing (many “Delta passengers” traveled on other carriers in the course of their trip). Similarly, American had a largely hub-and-spoke service topology, but had very little

code-sharing. Southwest’s network, however, was comprised of weaker hubs and had virtually no code-sharing.

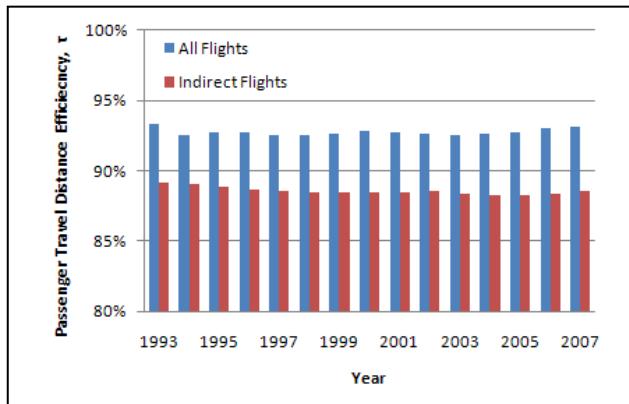


Figure 1: Variation in average τ over time for all airlines.

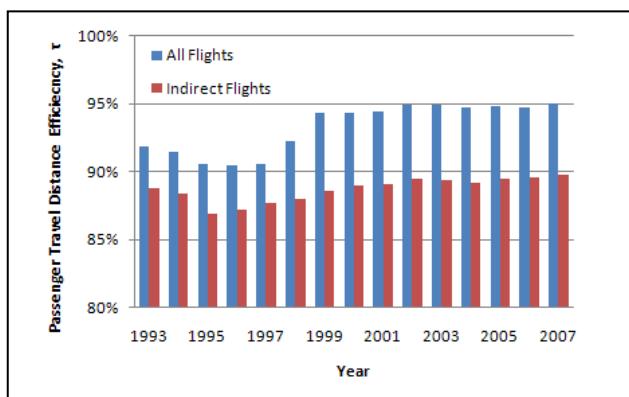


Figure 2: Variation in average τ over time for Southwest Airlines.

IV. NETWORK TOPOLOGY TRADE-OFF STUDY DESCRIPTION

A. Overview

The previous section investigated the historical trends of passenger travel distance efficiency for the ATS and selected airlines. In this section, the correlation between airline service route network configurations and passenger centered ATS efficiency is explored. The airline service route network topology examined here is on an annual scale and scheduling of actual flights is not considered. In other words, links in the service route network are simply paths which allow transporting of passenger from their origin to destination airports (nodes) for a particular annual demand. Further, all the airlines service routes are aggregated into one single network unless otherwise noted.

Different types of networks are generated under the topology generator discussed in the following section. For each topology type studied, the passenger travel distance efficiency (τ) as well as the number of connections required to transport the passenger on the shortest routes are calculated. Since scheduling is not considered, total passenger travel time is not explored but combining analysis from τ and the required

number of connections, travel time can be estimated under a set of assumptions.

Data on historical passenger demand and airport operations are extracted from the 2005 DB1B Survey and T-100 Domestic Segment data respectively, both available from the Bureau of Transportation Statistics [12]. The number of nodes in the network is kept constant at 304, representing the airports only in the continental US.

B. Network Topology Generator

Currently, random and scale-free networks are the most discussed types of network topology used for analysis. Scale-free networks are similar to the hub-and-spoke networks of the ATS where few nodes with high degree (i.e., number of links) maintain the connectivity throughout the network. Similar structure is also seen in protein networks, social networks and the World Wide Web [13]. The prime benefits of this structure are that all nodes are connected via relatively few links, and new nodes can be easily integrated as long as the hub nodes are functional. On the other hand, the main drawback of a scale-free network is that as the hub nodes become larger, the risk of a single point of failure increases significantly. Scale-free networks can be constructed using the Barabási-Albert (BA) model [13] which runs under the precept of a preferential attachment behavior where nodes with higher importance are granted a higher probability to attain a new link. In the BA model, importance of a node is valued by its local degree compared to the total degree of the network. In another words, the probability of node A linking with any other node B is

$$P_{\text{connect}}(A, B) = \frac{k_A}{\sum_{i=1}^j k_i} \quad (2)$$

where j is the total number of nodes in the network and k is nodal degree. For random networks, links between nodes are constructed based on a uniform probability distribution function which remains constant for all node pairs that may form a link. While random networks require more links for equal shortest-path connectivity compared to a scale-free, the single point of failure risk is much lower since all nodes are almost equally important in terms of the number of connections [13].

The Network Topology Generator (NTG) constructs a network with varying mix ratio of scale-free and random characteristics, based on the user input. The NTG algorithm first generates two networks, random and scale-free with equal total degree for the same node set. The NTG then arbitrarily selects links from the scale-free and random network and places it in the final network; the number of links chosen from either the scale-free or random network depends on the user input mix ratio mentioned earlier. For example, if the mixture ratio was 80% scale free, the NTG will chose 80% of the final links from the scale-free network generated in the initial step, while extracting the remainder 20% from the random network. Networks of different scale-free and random topology mix ratios will be examined for the impact on passenger travel

distance efficiency and number of connections required to fulfill demand.

V. RESULTS

A. Passenger Centered Efficiency

Topologies with six scale-free / random mix ratios and four different network densities were created for this study. Network density is the ratio between the total network degree and the number of possible links that can exist in a particular network size. Using the 6% density (2612 links) which was observed in the historical 2005 ATS network as a baseline, networks with 12%, 3% and 1% density were considered.

Figure 3 displays the τ for each topology type and Figures 4 through 7 show the ratio for number of connections required to fulfill the annual passenger travel demand. Each column shows the different network mix ratios. For example, "BA80" means 80% of the links came from the BA (i.e. scale-free) logic, while the remaining 20% is from the random network logic. In some cases, not all passengers can be transported from their origin and destination demand by available routes for the networks generated by the NTG. This is due to non-connected 'island' clusters that occasionally form in random networks with low density, but as shown in Table 2, only a small portion of the passengers in the 1% density topology fall under this category. In addition, results displayed in Figure 3-7 are an average value over 10 runs, and the fluctuation between each run is relatively small (<5% on average).

As expected, the travel distance efficiency increases for topologies with higher density and more scale-free characteristics (Figure 3). However, the difference in τ was considerably small between the higher and lower density networks. For example, τ in a network with 1% density was 37% less compared to a 12% density network under the BA100 mix ratio. In a network with 304 nodes, this 10% difference in density is equivalent to approximately 5000 links. Since most demand is still satisfied (as shown in Table 2) the service route network with 1% density was able to transport the same amount of demand with about 5000 fewer links, in exchange for lower travel distance efficiency. Further analysis between degree and travel distance efficiency may be a useful study for future ATS transformation efforts if links are considered as resources in constructing a network. However, higher network density significantly decreases the minimum number of connections required on the shortest distance route as it can be seen in Figures 4-7.

TABLE 2. PERCENT OF PASSENGER DEMAND THAT CANNOT BE SERVED

Network Density	NTG Topology Mix Ratio					
	BA 100	BA 80	BA 60	BA 40	BA 20	BA 0
12%	0%	0%	0%	0%	0%	0%
6%	0%	0%	0%	0%	0%	0%
3%	0%	0%	0%	0%	0%	0%
1%	0%	0.9%	1.6%	0.6%	1.6%	2.6%

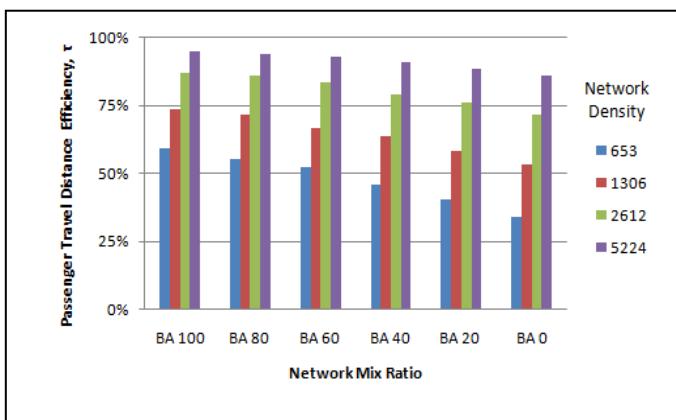


Figure 3. Passenger travel distance efficiency for different network mix ratio and network density.

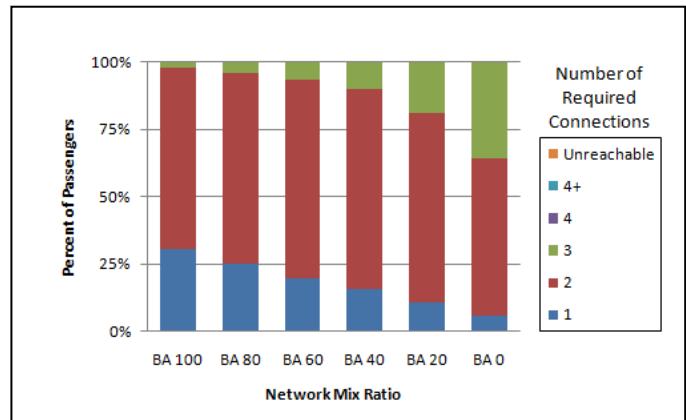


Figure 6. Number of required connections for network with 6% density.

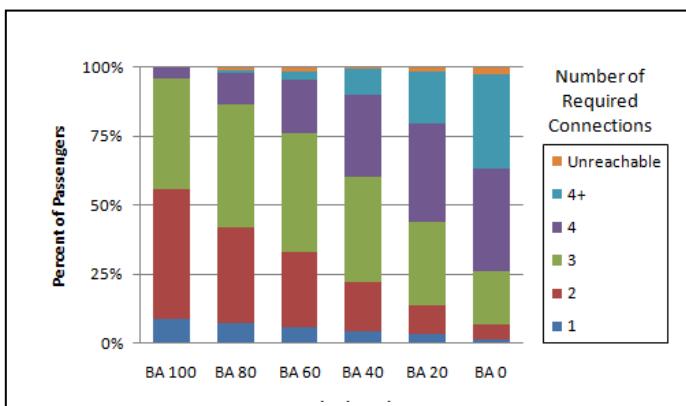


Figure 4. Number of required connections for network with 1% density.

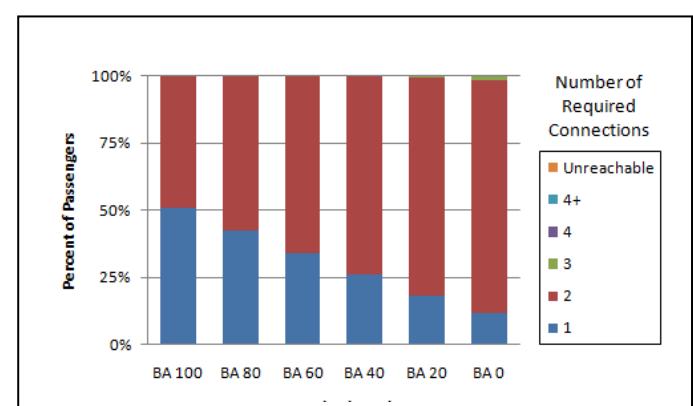


Figure 7. Number of required connections for network with 12% density.

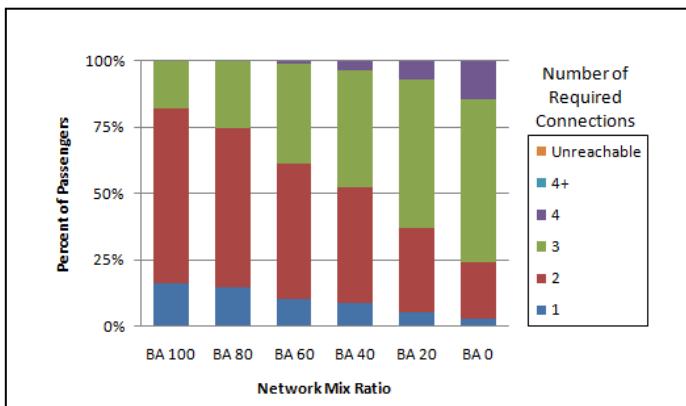


Figure 5. Number of required connections for network with 3% density.

B. Network Topology Robustness

Beyond the passenger centered efficiency discussed in the previous section, performance metrics regarding airline network robustness were also investigated. However, one cannot speak generally about robustness; instead, a class of possible disturbances must be specified in order to measure or estimate a particular robustness characteristic of the system. In terms of networks, there are two general types of “attack” that may cause disturbances: targeted and random. These attacks disable the function of a node (airport) and either temporarily or permanently remove it from the entire network, along with any associated links. Random attacks are arbitrary failures that can occur to any nodes within the network under certain probability; they usually represent incidents such as weather, accidents, and aircraft malfunctions. Targeted attacks, on the other hand, are failure of specific nodes which are usually due to an artificial cause. In the real world, targeted attacks may occur as terrorism, strike, or war-related issues.

Robustness of each network topology configuration is examined by measuring the degradation in τ and percent of passengers unable to travel after certain nodes are removed, mimicking targeted and random attacks. For targeted attacks, nodes with the highest degree are removed while for the random attack, nodes are removed randomly for the network. For each attack type on the different network configurations, five, ten and fifteen nodes were removed to observe how increasing number of failed nodes degrade the overall network performance. Tables 3 and 4 display the amount of performance degradation of the networks after the disruptions in terms of τ and percent of passengers unable to travel, respectively.

While both scale-free and random networks are fairly resistant towards random attacks, it can be observed that scale-free networks are extremely fragile towards targeted attacks until a certain network density is attained. Further, although the majority of passengers were unable to travel after targeted attacks on networks that exhibit the slightest scale-free characteristics, a fully random network is able to maintain routes to travel approximately 90% of the passengers.

TABLE 3. PRECENT REDUCTION IN PASSENGER TRAVEL DISTANCE EFFICIENCY (τ) AFTER DISRUPTION

Network Density	Disruption Type	Disabled Nodes	BA100	BA 60	BA40	BA0
3% (1306 links)	Random	5	0.06	0.41	0.51	0.48
		10	1.12	0.34	0.59	0.87
		15	0.01	0.37	0.98	1.18
	Targeted	5	9.99	8.00	6.70	1.48
		10	19.18	13.46	10.99	2.33
		15	25.10	17.23	13.94	3.27
6% (2612 links)	Random	5	0.02	0.33	0.13	0.18
		10	0.19	0.07	0.46	0.49
		15	0.38	0.62	1.10	0.93
	Targeted	5	2.88	3.27	2.29	1.01
		10	6.96	6.74	5.47	1.67
		15	12.40	11.04	8.68	2.48
12% (5224 links)	Random	5	0.08	0.18	0.29	0.37
		10	0.21	0.07	0.20	0.40
		15	0.11	0.31	0.20	0.68
	Targeted	5	1.11	1.02	1.12	0.36
		10	2.07	2.31	1.89	0.75
		15	3.65	3.51	3.12	1.07

TABLE 4. PERCENT OF DEMAND THAT CANNOT BE SERVED AFTER DISRUPTION

Network Density	Disruption Type	Disabled Nodes	BA100	BA 60	BA40	BA0
3% (1306 links)	Random	5	3.1	3.6	3.8	4.0
		10	7.3	6.4	7.8	8.2
		15	7.1	10.0	10.3	7.8
	Targeted	5	20.3	21.2	20.9	3.0
		10	36.5	37.6	38.5	8.1
		15	49.3	50.3	47.6	11.4
6% (2612 links)	Random	5	2.1	3.2	2.6	3.1
		10	5.6	6.2	5.0	5.6
		15	12.3	10.6	10.3	8.7
	Targeted	5	17.6	20.5	15.9	5.1
		10	34.6	34.7	33.4	8.5
		15	49.1	48.5	48.2	13.0
12% (5224 links)	Random	5	3.6	4.2	4.3	3.3
		10	7.8	5.1	6.2	7.9
		15	9.0	9.5	7.0	10.2
	Targeted	5	17.5	17.4	16.4	2.1
		10	32.0	33.2	32.2	5.5
		15	44.4	46.0	44.5	8.1

In summary, what is meant by a “favorable” network configuration for the ATS is quite different depending on the focus of the efficiency metric. From the perspective of τ and number of connection required to transport passengers under historical patterns, a network that shows strong scale-free characteristics seems more suitable. However, a random configuration seems to be more ideal from a robustness standpoint, since they are more resistant to both targeted and random attacks compared to a scale-free type topology.

C. Topology of the Current ATS Network

The resemblance of the current ATS network with the various topologies generated by the NTG was examined using a degree distribution plot, shown in Figure 8. Each line shown is a linear regression of the degree distribution in log-log scale using the least-squares method for the BA100 and BA20 network generated over 10 runs (besides the historical data).

All networks, including the 2005 data have a total degree of 2612 (6% density). It was observed that the actual ATS network resembled a network with more random characteristic than scale-free in terms of the regression line slope, an unexpected result given that many airlines tend to use the hub-and-spoke business model, which should show behavior similar to scale-free. This may have been an artifact of the reduction of network size (BTS tracks more than 2500 airports, although most are usually inactive), or the aggregation of multiple airlines. However, since network science is a relatively new area of study, a formal network comparison methodology does not currently exist. Further research to quantify network topology traits is necessary to better utilize the research efforts pursued by the authors.

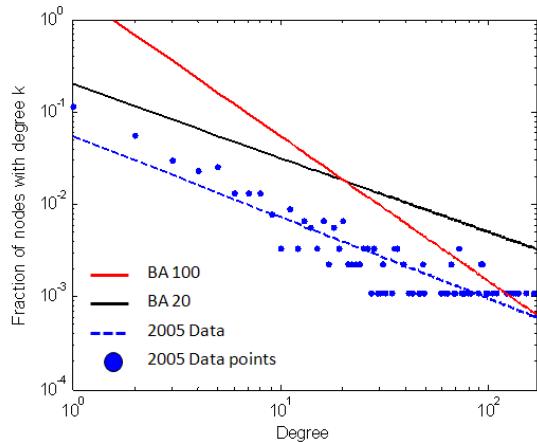


Figure 8. Degree distribution of historical and NTG networks.

D. Comprehensive Analysis

Figure 9 shows an example of collectively analyzing the data generated in the previous section in order to understand some of the holistic trade-offs in architecting the future ATS network. Increasing the scale-free characteristics (the BA value for the NTG) in a low density network significantly increases τ , but also increases the potential damage from a targeted disruption at the same time. Further, the effectiveness of scale-free characteristic for a high τ declines after the network reaches a certain density level. From this particular analysis, decision-makers of the ATS would be able to know the minimum level of network density that needs to be attained in order to achieve a certain level of resistance to failure modes while also keeping τ at an acceptable level for the passengers. Similar trade-off studies that correlate individual aircraft performance and fleet mix to fuel efficiency and τ are currently being developed.

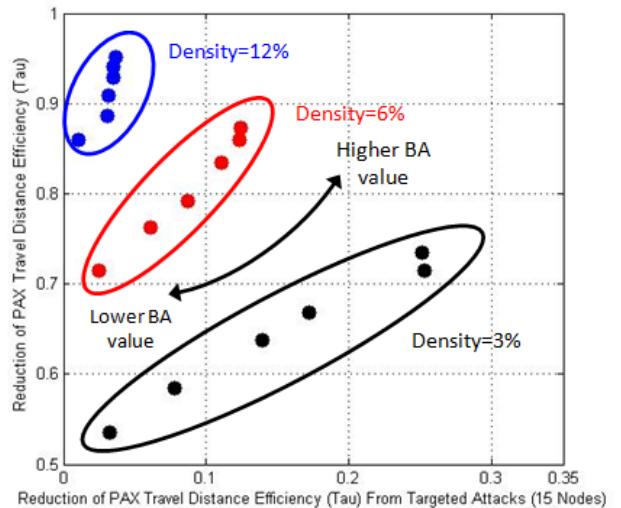


Figure 9. Trade-off between τ and network robustness against targeted attacks.

VI. CONCLUSION AND FUTURE WORK

Research reported in this paper provided an initial investigation on how system configurations for large scale systems like the ATS may differ depending on stakeholder viewpoints. A rudimentary trade-off among different airline service network configuration was examined for both efficiency in processing travel demand and resistance to various failure modes. Current results presented throughout the paper show that the favorable network configurations may lie on opposite extremes depending on the different objectives examined. We do recognize that the control of the actual service route network structure is distributed among the various airlines; there is no central route-allocating architect. However, the results reported here provide quantitative bounds on the efficiency and robustness of different network configurations that could serve as targets for system transformation. Given these targets, policymaking bodies, as well as airline enterprises, can use the influence factors they do control to drive overall system behavior towards these preferred network configurations. Before further work in ATS transformation is commenced, objectives need to be prioritized in order to clarify the ideal configuration of the future ATS.

The results reported in this paper describe only the initial investigation of ATS architecture trade-offs; as such, there is more work to be done. The first step is to extensively review and construct efficiency metrics that can represent how well a particular stakeholder's objectives are met under various architecture configurations. The study on τ , number of required connections, and disruption resistance studied in this paper mainly involve the passenger, airlines, and regulators, but the actual ATS involves additional stakeholders that need to be considered such as airports and air traffic controllers. The second step is to construct a series of analytical methods that can cross multiple timescales since each stakeholder's objective may reside within a different timescale. For example,

airports often use arrival and departure operations that can be processed per minute or per hour. However, for stakeholders that emphasize long-term ATS capabilities such as sustainability, trade-offs cannot be made under the current approach. In the short term, we plan on expanding the boundaries of this study to aircraft and fleet mix design, which would also incorporate efficiency on fuel use towards different network configuration options.

ACKNOWLEDGMENT

The authors acknowledge sponsorship of this research by the FAA Air Traffic Organization Office of Performance Analysis and Strategy. The contributions of Sricharan Ayyalasomayajula at Purdue University are also acknowledged and appreciated.

REFERENCES

- [1] DeLaurentis, D.A., "Understanding Transportation as a System-of-Systems Design, Problem," *AIAA Aerospace Sciences Meeting and Exhibit*, 10-13 Jan. 2005.
- [2] R. Babikian, S. P. Lukachko and I. A. Waitz, The historical fuel efficiency characteristics of regional aircraft from technological, operational, and cost perspectives. *J. Air Trans. Mgmt.* Fairfax, VA, vol 8, issue 6, pp. 389-400, November 2002.
- [3] B. Manley and L. Sherry, The impact of Ground Delay Program (GDP) rationing rules on passenger and airline equity. *Integrated Communications Navigation and Surveillance (ICNS) Conference*, Bethesda, MD, May 2008.
- [4] G. Calderón-Meza, L. Sherry, and M. Consiglio, Multi-agent simulation of NAS infrastructure utilization in the presence of airline policies. *Integrated Communications Navigation and Surveillance (ICNS) Conference*, Bethesda, MD, May 2009.
- [5] I. J. Tetzloff and W. A. Crossley, An allocation approach to investigate new aircraft concepts and technologies on fleet-level metrics. *9th AIAA Aviation Technology, Integration and Operations (ATIO) Conference and Aircraft Noise and Emissions Reduction Symposium (ANERS)*, Hilton Head, SC, 21-23 Sep. 2009.
- [6] Meyn L, Romer T, Roth K, Bjarke L. Preliminary Assessment of Future Operational Concepts Using the Airspace Concept Evaluation System. *Proceedings of AIAA Aviation Technology, Integration and Operations Forum*, Chicago, Illinois, AIAA-2004-6508, 2004.
- [7] Reynolds, T., Barrett, S., Dray, L., Evans, A., Kohler, M., Morales, M., Scafer, A., Wadud, Z., Britter, R., Hallam, H., Hunsley, R., 2007. Modeling Environmental & Economic Impacts of Aviation: Introducing the Aviation Integrated Modeling Project. *AIAA Aircraft Technology, Integration and Operations (ATIO) Conference*, Belfast, Northern Ireland, 18-20 September 2007.
- [8] Guimera, R., Mossa, S., Turtschi, A., Amaral, L., 2005. "The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proc. Nat. Acad. Sci. U.S.A.*, 102(22), pp. 7794-7799.
- [9] Bonnefoy, P., Hansman, R., 2007. Potential Impacts of Very Light Jets in the National Airspace System. *Journal of Aircraft*, 44(4), pp. 1318-1326.
- [10] DeLaurentis, D., Han, E-P., Kotegawa, T., 2008. Network-Theoretic Approach for Analyzing Connectivity in Air Transportation Capacity Networks. *American Institute of Aeronautics and Astronautics Journal of Aircraft*, 45(5), pp. 1669-1679.
- [11] Kotegawa, T., Han, S., DeLaurentis, D., Implementation of Enhanced Network Restructuring Algorithms and Scenarios for Improved ATO Forecasts, *9th AIAA Aviation Technology, Integration and Operations (ATIO) Conference and Aircraft Noise and Emissions Reduction Symposium (ANERS)*, Hilton Head, SC, 21-23 Sep. 2009.
- [12] "Transtats," *Bureau of Transportation Statistics (BTS)*, accessed on 1 Dec. 2009, from <http://www.transtats.bts.gov/homepage.asp>
- [13] Albert, R., Barabási, A., 2002. Statistical mechanics of complex networks. *Rev. Mod. Phys* 74, pp. 47-97.

Estimation and Comparison of the Impact of Single Airport Delay to the National Airspace System using Multivariate Simultaneous Models

Yu Zhang & Nagesh Nayak

Department of Civil and Environmental Engineering
University of South Florida
Tampa, Florida, USA
yuzhang@eng.usf.edu

Tony Diana

Federal Aviation Administration
Washington DC, USA
tony.diana@faa.gov

Abstract— The U.S. air transport as we all know is under significant stress with frequent delays and congestion. Airports are considered as bottlenecks of the National Airspace System (NAS). The major causal factors of flight delay at one airport are over-scheduling, en-route convective weather, reduced ceiling and visibility around airports, and upstream delay propagation. Meanwhile, the delay occurred at this airport will be passed on to other airports in the NAS. Hence, to optimally allocating resource for airport capacity expansion, it needs to quantify the impact of single airport delay to the NAS and vice versa. This research explores the methodology to analyze not only airport delay impact to the NAS, also explore if the delay spillover is widely dispersed across 34 OEP airports or more concentrated using multivariate simultaneous regression models. Three stage least square (3SLS) is used to regress the models and obtain coefficients for the multivariate equations.

Keywords-Airport delay; NAS delay; delay propagation; 3SLS

I. INTRODUCTION

Airport congestion and delay has been the focus of intense research since last few decades. Many major airports in U.S. have significant delay problems due to increased air passenger demand. According to the Department of Transportation's Bureau of Transportation Statistics (BTS) only 79.10% of arrivals were on time from October 2008 to October 2009 [1]. The causes of flight delays include air carrier delay, late arriving, the National Airspace System (NAS), security, and extreme weather. Among these causes, the delays due to aircraft arriving late account for more than 30 percent of total flight delays. As a result of the network structure of the NAS, delay at one airport is likely to affect delays at other airports.

The NAS is a complex system comprising of a large number of airports. It is affected by unexpected events such as adverse weather, equipment outages, aircraft maintenance problem, airline crew issues, and others. All these factors make the NAS a complex and stochastic system. The Next Generation Air Transportation System (NextGen) envisions a highly efficient NAS by 2018 [2] when the total flight delay will be reduced by 30 to 40 percent in comparison to a do-nothing scenario. There are a number of ways that need to be

explored and implemented before achieving such a goal: adding or extending runways, developing innovative technologies and procedures, etc. All these alternatives require enormous capital investment. One of the five-year plans that regulates the NAS modernization projects, known as Federal Aviation Administration's Capital Investment Plan (CIP), intended to invest about \$16.6 billion from 2010 to 2014 in projects that modernize the existing system, increase airspace capacity, and introduce new technologies to achieve the planned NextGen capabilities [3]. Considering the airport capacity expansion, for optimally allocating resource, there is a need to quantify, not only the local benefits of expansion, but also the advantages of the expansion to the system. From an air transportation planning and policy point of view, sufficient tools are needed to test the system-wide effects of such investment activities and help further strategic planning.

Various researchers have tried to understand the microscopic perspective of delay propagation (Beatty et al. [4], Schaefer and Millner [5], Schaefer et al. [6] and Ahmad Beygi et al. [7]). Nevertheless, their studies capture details of only a few components of the NAS such as specific airports, sectors, or individual flights, but fail to reflect the system overall. A former research done by Zhang and Nayak [8], captures the delay propagation phenomena from a macroscopic point of view. It used multivariate simultaneous-equation regression model to study the impact of single airport delay to the system and vice versa [8]. Specifically, we applied our model to Chicago O' Hare International Airport (ORD) and LaGuardia Airport (LGA). These two airports have attracted enormous attention for significant and persistent delays. The research explored causal factors of the delays at these two airports and compared their system-wide impacts. The estimated results quantified the interdependency between flight delay at an individual airport and other 34 Operational Evolution Partnership (OEP) airports taken together as the NAS. Scenarios were also constructed to analyze how capacity improvements or new demand management strategies at those two airports would affect the performance of the rest of the NAS.

This research presents a macro perspective and proposes not only to investigate the impact of single airport delay to the NAS, but also to explore how the delay spillovers is widely dispersed across the (OEP) 34 airports (see Appendix). Causal factors of average daily arrival delays are explored and multivariate equations are developed for all the airports under consideration along with the NAS. The average daily-arrival delay is the dependent variable in the equation for each airport and the NAS, while it is also taken as an independent variable in the equations of other airports and the NAS. The estimated coefficients can be interpreted as the marginal effect of delay increase of that airport to the other airports or the NAS. This type of model is widely used in economics and business management research studies. We can use the three stage least square (3SLS) method to regress the model.

The remainder of the paper is organized as follows: Section 2 summarizes existing literature on delay propagation and discusses factors affecting delay. Section 3 specifies multivariate simultaneous equations and 3SLS method. Section 4 presents a summary of the results. Section 5 concludes the study and provides suggestions for future research studies.

II. LITERATURE REVIEW

Beatty et al. [4] developed the concept of a delay multiplier for understanding the effect of initial flight delay on an airline's operating schedule. They assumed that various airline resources such as crew members, aircraft, passengers, and gate space affect flight delay. The delay multiplier was used to determine all potential downstream flight delays connected to that initial flight. Their research concludes that the existence of a delay multiplier is due to the branching nature of crew and aircraft sequences. The research estimated the delay propagation from one airport to the other based on the connectivity of airline's operating resources and its schedule.

Delay propagation has also been studied by Schaefer and Millner [5] using the detailed policy assessment tool. They modeled the propagation of delay throughout airports and airspace sectors given inputs such as air traffic demand and airport capacities. They synthesized aircraft assignment given the air traffic data from Official Airline Guide (OAG) and then used the information to simulate delay propagation according to departure and arrival queues between origin and destination airports. Three airports were analyzed using several combinations of Visual Meteorological Conditions (VMC) and Instrument Meteorological Conditions (IMC) when capacities reduced due to inclement weather. The results show that the delay augments with prolonged duration of IMC at the airports. They also concluded that although the propagation effect for the first leg was significant, it diminished along each subsequent leg.

Further research by Schaefer et al. [6] developed an analytical model to separate controllable factors that influence delays and their propagation in the NAS from other factors that are random variables in a given scenario. The controllable factors are scheduled and minimum airport turnaround time, slack for airport turnaround time, scheduled and minimum flight time between airports, and fixed flight time allowance,

while the variable factors considered in the research were variable airport turnaround time and variable airport flight time. The model analyzed the interaction between fixed and variable delay components at each airport under both VMC and IMC conditions and emphasized the importance of schedule parameters on delay propagation in the NAS. Their study shows that airports with less slack time between flights had more delay.

A recent research by Ahmad Beygi et al. [7] explores a similar observation in terms of slack time between two flights. Their study indicates that the delay of one flight can propagate to disrupt one or many subsequent downstream flights that await the aircraft and crew from the delayed flight. In such case, the presence of well-planned slack between flights is critical for absorbing the disruption.

The studies discussed above attempt to show how common resources and weighted airline schedules can be major causes of delay propagation. These research studies are clear indicators that the issue of delay propagation at airports is prevalent.

A macroscopic research by Diana [9] proposed a methodology to compute delay propagation from airports based on the Discrete Fourier Transform (DFT). The airports sampled in his study vary in terms of location and traffic throughput. The research assumed that the delay propagation is similar as wave propagation where the delays represent signals and the NAS acts as the medium. Airlines anticipate delays and build precautionary buffer in their schedule to absorb the propagation effects. In his study, he applied the delay concept in airline on-time performance, i.e. only arrival flights with more than fifteen minutes delay past schedule are considered as delayed flights. Diana tried to investigate whether market concentrated airports (i.e. with higher traffic throughput) have more delay propagation effects than less concentrated airports. The outcomes shows that, when delay propagation is considered as a signal through the system, it is not dependent on the degree of market concentration.

A recent study done by Laskey et al. [10] takes into consideration the dynamic aspects of flight delay, such as weather effects, wind speed, flight cancellations, and others, to estimate delay propagation in the NAS. They used Bayesian Networks (BN) to quantitatively analyze major factors affecting each delay component and the relationship among the delay components. In their study, flight arrival delay was decomposed into Gate-In Delay, Turn Around Delay, Gate-Out Delay, Taxi-Out Delay, Airborne Delay, and Taxi-In Delay, each of which was considered as a dependent variable for that phase of the flight, with delays from previous phases as independent variables. The principal objective of this research was to estimate the impact of changes in tactical decisions and policies with respect to the ground delay program (GDP), rescheduling, and cancelled flights on delay in the system. Nevertheless, only three months of data were used to identify the critical phase of the flights from ORD and Hartsfield-Jackson Atlanta International Airport (ATL).

Hansen and Zhang [11] devised a macroscopic technique to study the delay propagation in the NAS. They studied the operational performance at LGA under different demand

management regimes using multivariate simultaneous equation regression model. The outcome of that research shows that, according to historical data from 2000 to 2004, the increase in one minute average-daily-arrival delay at the LaGuardia when compared to airline schedule causes an increase in the average-daily-arrival delay at non-LGA airports by 1.7 minute [4]. The research identified various factors causing arrival delay at LGA and non-LGA airports and estimated the impact of each of these factors on the total delay.

Our study seeks to extend our previous research, as mentioned in the Introduction, by estimating the interaction between flight delay at one single airport and delay at the other 34 OEP airports and the rest of the NAS. This study quantifies the performance improvement due to capacity expansion and demand management strategies in terms of reducing congestion and delay while controlling for other factors.

III. METHODOLOGY

Multivariate simultaneous equation regression model is a form of statistical model with a set of multivariate equations where the dependent variable in one equation could be independent variable in other equations. In addition, the error terms in the equations could be correlated. This type of model is widely used in economics and business management research studies. In our study, multivariate simultaneous equations are generated for each of the 34 OEP airports excluding Honolulu International Airport (HNL). Additionally, a separate equation is included for the delay in the rest of the NAS by combining all the remaining ASPM77 airports together. As shown in Fig. 1, equations for a single airport share the similar set of independent variables while the NAS contains different variables. The error terms of all the equations are correlated to each other.

Three stage least square (3SLS) method can be used to regress the model and obtain coefficients for the multivariate equations. 3SLS combines two statistical techniques, one is the two stage least square (2SLS), and the other seemingly unrelated regression (SUR). In the first stage of 2SLS, each endogenous covariate in the equations of interest is regressed on all of the exogenous variables in the model, including both

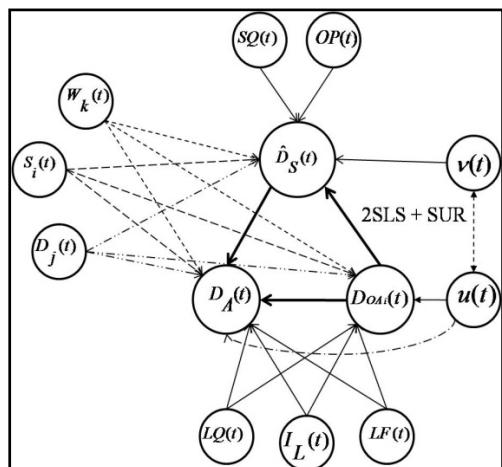


Figure 1. Interactions between a Single Airport and rest of the NAS

exogenous covariates in the equation of interest and the excluded instruments. The predicted values from these regressions are obtained. In the second stage, the coefficients in the equations of interest are estimated by regression, except that in this stage each endogenous covariate is replaced with the predicted values from the first stage. SUR is an extension of linear regression model allowing correlated errors between equations. It is a way of improving the efficiency of estimation equations jointly as it provides consistent estimates for linear regression models when explanatory variables are correlated with the error term.

A. Model variables

Most of the model variables are defined in an earlier paper however; we refined the explanatory variables given the new and extended dataset. The data used in this study is the Aviation System Performance Metric (ASPM) data at 77 airports from 2000 to 2008. For each OEP airport and the rest of the NAS, the average daily arrival delay is a function of average arrival delay at other airports, deterministic queuing delay caused by the over-scheduling or supply-demand imbalance due to capacity deficiency, adverse weather, and flight operations together with dummy variables indicating the seasonal and yearly effects.

- Average Daily Arrival Delay

Average daily arrival delay represents the dependent variable in our model. This delay is defined as scheduled daily arrival delay for all ASPM arrivals based on the Official Airline Guide (OAG). Only arrival delays are used as the delay metric, as it is observed that there is a high correlation between arrival and departure delay for both individual airports and the NAS.

- Deterministic Queuing Delay

Deterministic queuing delay indicates the operational demand and supply relationship at the airport. The arrival count is the actual number of arrivals at the airports in 15 minutes, which is restricted by the number of flights need to land and airport arrival rate (AAR) during the same time period. In another words, if the number of flights waiting to land is larger than the AAR rate, then the arrival count is the AAR rate, otherwise, the arrival count is the number of flights need to land.

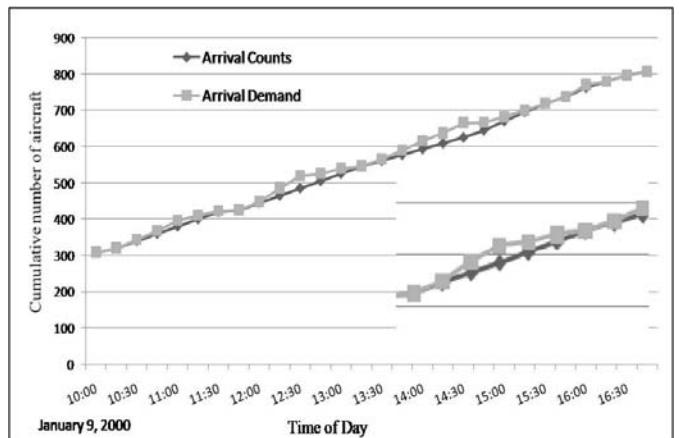


Figure 2. Queuing diagram of arrivals at ORD

The cumulative flight demand in a quarter hour is the remaining scheduled arrival demand until the end of the quarter hour [11]. Fig. 2 shows that the arrival count curve is always less than arrival demand since arrival counts are either restricted by arrival demand or the capacity of the system. The daily average queuing delay at an airport is calculated by dividing the area between the curves, which is known as total queuing delay, by the total number of arrivals at the airport for that day [11]. The same definition applies to the NAS model, considering arrivals at all the remaining ASPM77 airports together.

- Adverse Weather

Adverse weather has always been one of the important factors causing delay. In the NextGen environment, new technologies and procedures are being developed to mitigate poor weather conditions [2]. The model captures the adverse weather effects in two ways: convective weather index and IMC ratio. First, convective weather is integrated into the model by dividing the U.S.A. into regions of 10 degrees latitude by 10 degrees longitude. For each region, the proportion of weather stations reporting thunderstorms is obtained from the Surface Summary of Day database maintained by the National Oceanographic and Atmospheric Administration (NOAA). Thus, the convective weather index for a particular region is calculated as the ratio of the number of stations reporting thunderstorms by the total number of stations in the same region. Secondly, the IMC ratio is calculated as the proportion of the day in which the airport was under IMC conditions.

- Passenger Load Factor

Individual airport models includes monthly passenger load factor as one of the explanatory variable. It is the monthly average ratio of the number of passengers by the number of seats available at the airport under consideration. It is assumed that higher passenger load factor leads to longer average daily arrival delay since it causes uncertainty to smooth daily operations.

- Total Flight Operations

The NAS model also contains total flight operations as one of the variables. It captures the effects of total traffic volume on the delay in the NAS. This variable also accurately explains the congestion period in the system.

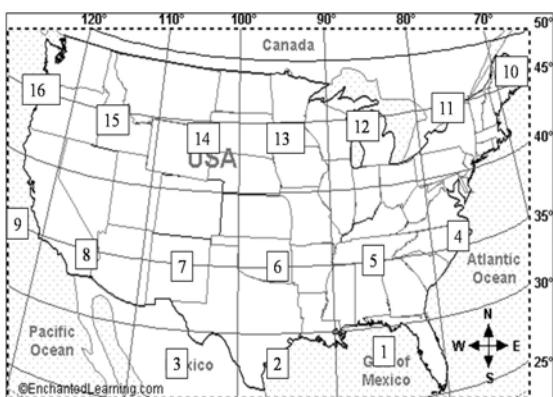


Figure 3. USA Weather Regions

- Seasonal and Yearly Dummy Variables

Dummy variables are introduced to indicate seasons and different years from 2000 to 2008 among which year 2001 has been divided as before and after 9/11 event.

B. Model 1 for an individual airport

The model for an individual airport decomposes average daily delay into components related to different delay causal factors. The explanatory variables include average arrival deterministic queuing delay, average observed arrival delay at other airports, adverse weather, seasonal effects, yearly dummy variable, passenger load factor, and the others.

$$D_A(t) = \alpha_A + \beta_1 D_S(t) + \beta_2 D_{OAi}(t) + \beta_3 LQ(t) + \beta_4 LQ^2(t) + \beta_5 LF(t) + \beta_6 IA(t) + \beta_7 IA^2(t) + \sum_k \lambda_{kA} W_K(t) + \sum_i \omega_{iA} S_i(t) + \sum_i \theta_{jA} D_j(t) + v(t) \quad (1)$$

C. Model 2 for the rest of the NAS

The model for the NAS decomposes average daily delay at rest of the airports that excludes 34 OEP airports. The explanatory variables include variable delays at individual airports, convective weather, total operations, seasonal effects, yearly dummy variable, and other factors.

$$D_S(t) = \alpha_S + \gamma_1 OP(t) + \gamma_2 D_{OAi}(t) + \gamma_3 SQ(t) + \sum_k \lambda_{kS} W_K(t) + \sum_i \omega_{iS} S_i(t) + \sum_i \theta_{jS} D_j(t) + u(t) \quad (2)$$

The notations in the above two models are described as follows:

$D_A(t)$ = Average observed arrival delay against schedule at individual airport on day t;

$D_{OAi}(t)$ = Average observed arrival delay against schedule at other individual airport (i) on day t;

$D_S(t)$ = Average observed arrival delay at airports other than LGA or ORD on day t;

$LQ(t)$ = Average arrival deterministic queuing delay at individual airport on day t;

$LF(t)$ = Passenger load factor at the airport on day t;

$IA(t)$ = Daily IMC ratio recorded at individual airport on day t;

$OP(t)$ = Total operations (arrivals) of the system on day t;

$SQ(t)$ = Weighted average arrival deterministic queuing delay of the system on day t;

$W_K(t)$ = Weather index of region k on day t;

$S_i(t)$ = Seasonal dummy variable, set to 1 if daily arrival delay is observed in quarter i and 0 otherwise;

$D_j(t)$ = Yearly Dummy Variable, set to 1 if daily arrival delay is observed in year j and 0 otherwise;

$v(t), u(t)$ = Stochastic error terms; and

$\alpha, \beta, \lambda, \omega, \theta$, and γ are coefficients.

IV. ESTIMATION RESULTS

The 3SLS method has been used to estimate the coefficients in the simultaneous equation models. The estimated coefficients for average queuing delay for most of the airports except PIT, MEM, SAN and TPA airports indicate that supply and demand imbalance is likely to be a major contributing factor to average daily arrival delays. However, the negative coefficient for the quadratic term of average queuing delay shows that this factor reduces as average queuing delay increases. This study explores the delay propagation from other airports and the rest of the NAS to an individual airport. The estimation results show that the other airports around the same geographical region or the other airports operating as a hub for the same carrier contribute significantly on the delay at the individual airport. For instance, the airports significantly affect the arrival delay at ATL are CLT, CVG, MEM, BWI and MCO, which are all located in the eastern part of the country. Similar regional phenomena can be observed and summarized in Table 1. Counter-intuitively, several airports have negative delay propagation effects on some other airports. For example, the delay increase at LGA will reduce the delay at JFK, MCO, STL, DTW, and CLT. The IMC ratio is likely to impact the delay at almost all the airports except PIT. Most of the airports are affected significantly by the convective weather index in the same region where they are located except BOS, CVG, LAS, MIA, PDX, SLC and SAN. It is also observed that a few airports like DEN, BWI and MEM are affected by thunderstorms occurring at destinations. In addition, convective weather at region 2 and 6 which represent southern states contribute considerably to delay at the rest of the NAS airports.

As long as the weather pattern is captured by convective weather index and IMC ratios, seasonal dummy variables in the model only reflect the seasonal difference of airline scheduling. The estimates for the seasonal effect show that their impact on delay is very small in comparison to other factors. Interestingly, for most of the airports, the winter seasonal effect shows highest amount of delay as compared to other seasons. However for the airports in the southern parts of the country like MCO, FLL, ATL, TPA and LAS, delays are higher during spring. The results from yearly dummy Variables have a large impact on average daily arrival delay. The estimated coefficients for the dummy variables provide a better perspective on how delays vary in comparison to different time periods. According to FAA, 34 OEP airports are categorized into different regions (different from the convective weather regions that we have defined earlier) [12]. The trends of average arrival delay for all the airports along with the NAS are shown in Fig. 4 to 11. Fig. 4 shows that the average arrival delays at all the airports in ASO region, except MEM and MIA, decreased from 2000 to 2005 but then increased in 2007. At MIA, average daily arrival delay increased continuously from 2000 to 2008. In Region AWP, as shown in Fig. 5, the delay at LAX and SFO decreased drastically after 9/11 and slowly approached the level of pre 9/11 in 2006. For LAS and PHX in the same region, however, the delay increased immediately after 9/11. Fig. 6 shows the delay trends of the airports in ANM region, which comprises

of airports in the north-west of the country. The average arrival delay at those airports was higher in 2007, but still lower than the pre 9/11 level.

The north-central part of the U.S. is represented by AGL region (Fig. 7), which consists of many connecting airports for east-west air traffic. The arrival delay at most of the airports reduced after 9/11 and then increased gradually afterwards. Nevertheless, the delay at MDW airport has significantly reduced from 2000 to 2008, except a rise-up in 2006. The ASW region (Fig. 8) consisting of airports from Texas state had arrival delay reaching its peak in year 2007-08. The north-eastern part of the country that has a few of the world's busiest airports is represented by AEA region (Fig. 9). This region consists of the largest number of airports as compared to other regions. For all the airports, except IAD, the average arrival delay reduced after 9/11, slowly increasing thereafter and reaching its peak in 2007. The average arrival delay at rest of the airports (Fig. 10 and Fig. 11) reduced after 9/11 and reached its peak in 2007.

V. CONCLUSION

Airport delay has always been a major problem for the aviation industry. Most previous studies estimate the delay propagated through an individual flight from an airport to the system. This research illustrated the effectiveness of applying multivariate simultaneous equation model to study delay propagation from a single airport to other airports and to the rest of the system, and vice versa. The model developed for airports takes into account all the delay causal factors mentioned earlier and can include more in future models. The model estimates the effect of each of these factors using the 3SLS method. This method is generally used to deal with the bidirectional relationship that exists between dependent and independent variables and suitable for the equations with correlated error terms. The estimated results help quantify the interdependency between flight delays at different airports and the NAS.

The regression results show that queuing delay and adverse weather are major delay causal factors at most of the studied airports. Passenger load factor is an important factor at some of the hub airports like MDW and MEM but not others. Airports located in same geographical regions had more interactions than others. Major airports like ATL, ORD, PHX and EWR had more impact on average arrival delay than other airports. BOS, MIA and BWI had least impact on arrival delay at other airports. The graphical representation for different time periods from the year 2000 to 2008 demonstrates the significantly delay variation. Most of the airports, with a few exceptions, had their delay reduced after 9/11 and gradually increased back to pre 9/11 level with a peak in 2007.

As the next step of this research, we are exploring more explanatory variables such as capacity ratio, runway configurations, wind speed, demand management programs for all the airports and conduct more experiments on the specification of the model. To improve the efficiency of the model we also need to check the availability of some of the surrogates for our existing variables like passenger load factor, IMC ratio, etc. We plan to look into different delay definitions as well. Depends on the implementation of the model, arrival

delay could be measured according to airline schedule or flight plan. We also need to find out the causes for delay at each specific individual airport. We would also like to explore how

the delay in regional airport system affects other airports and the rest of the NAS. A good example will be the New York regional airport system containing LGA, EWR, and JFK.

TABLE I. Interactions between Different Individual Airports and the NAS

Airports	Airports Contributing to Average Arrival Delay	Airports Reducing Average Arrival Delay
ATL	CLT (0.264), CVG (0.220), MEM (0.260), BWI (0.160), MCO (0.229), NAS(0.324)	MIA (-0.177)
BOS	ATL (0.051), CLT (0.262), CVG (0.218), MEM (0.249), NAS (0.302)	MIA (-0.182)
BWI	ATL (0.042), EWR (0.131), PHL (0.094), IAD (0.093)	
CLE	DTW (0.115), EWR (0.088), PIT (0.146), NAS (0.472)	
CLT	ATL (0.070), PHL (0.107), PIT (0.148), NAS (0.3220)	LGA (-0.086)
CVG	ATL (0.051), ORD (0.042), PIT (0.176), NAS (0.235)	
DCA	ATL (0.038), IAD (0.142), PHL (0.154), NAS (0.195)	
DEN	ORD (0.039), PDX (0.228), PHX (0.159), SLC (0.162), DTW(0.090)	BOS (-0.009), CLE (-0.125)
DFW	IAH (0.087), PHX (0.145), NAS (0.224)	BOS (-0.019)
DTW	EWR (0.080), FLL (0.141), IAD (0.131), ORD (0.060), NAS (0.228)	BOS (-0.014), BWI (-0.174), LGA (-0.087)
EWR	CLT (0.248) and NAS (1.210)	MSP (-0.089), SAN (-0.354)
FLL	EWR (0.108), MCO (0.413), PHL (0.111), TPA (0.355)	
IAH	NAS (0.291)	
IAD	EWR (0.088), PHL(0.094), NAS (0.479)	SAN (-0.265), DFW (-0.066)
JFK	BOS (0.050), EWR (0.277), FLL (0.198)	LGA (-0.130)
LAS	DEN (0.084), LAX (0.114), PHX (0.233), SFO (0.055), SLC (0.100)	BOS (-0.016)
LAX	LAS (0.139), MEM (0.110), PHX (0.129), SFO (0.106), SLC (0.083)	BOS (-0.013)
LGA	EWR (0.385) and NAS (1.574)	BOS (-0.094)
MDW	DTW (0.188), ORD (0.264), PHL (0.089), NAS (0.341)	
MEM	ATL (0.053), CVG (0.149), MSP (0.072), ORD (0.037), NAS (0.383)	BWI (-0.149)
MIA	EWR (0.068), FLL (0.197), MCO (0.275), TPA (0.210)	
MSP	DTW (0.118), ORD (0.041), SLC(0.121), NAS (0.216)	
ORD	DTW (0.331), MDW (0.843), MSP (0.238), NAS (0.562)	BWI (-0.326)
PDX	DEN (0.456), LAS (0.049), SEA (0.292), SFO (0.056), SLC (0.117)	
PIT	DTW (0.317), MDW (0.781), MSP (0.219), ORD (0.425)	
PHL	CLT (0.22), EWR (0.099) and NAS (0.661)	
PHX	DEN (0.077), LAS (0.107), SLC (0.087)	BOS (-0.013), PDX (-0.177)
SAN	EWR (0.052), LAS (0.217), LAX (0.177), PHX (0.162), SFO (0.089), SLC (0.091), STL (0.043)	BOS (-0.014), BWI (-0.105)
SEA	FLL (0.091), PDX (0.599)	BOS (-0.011)
SFO	EWR (0.129) and NAS (0.393)	BOS (-0.090)
SLC	DEN (0.096), FLL (0.108), PDX (0.268), PHX (0.091), SFO (0.034)	BOS (-0.016)
STL	EWR (0.105), ORD (0.084), PHX (0.115) and NAS (0.170)	BOS (-0.025), LGA (-0.094)
TPA	ATL (0.069), CVG (0.095), EWR (0.071), FLL (0.131), MCO (0.146), PHL (0.075)	BOS (0.010)
NAS (System)	ATL (0.031), CVG (0.068), EWR (0.065), LAS (0.050), MEM (0.113), ORD (0.029), PHX (0.100), SLC (0.059), STL (0.043)	BOS (-0.006)

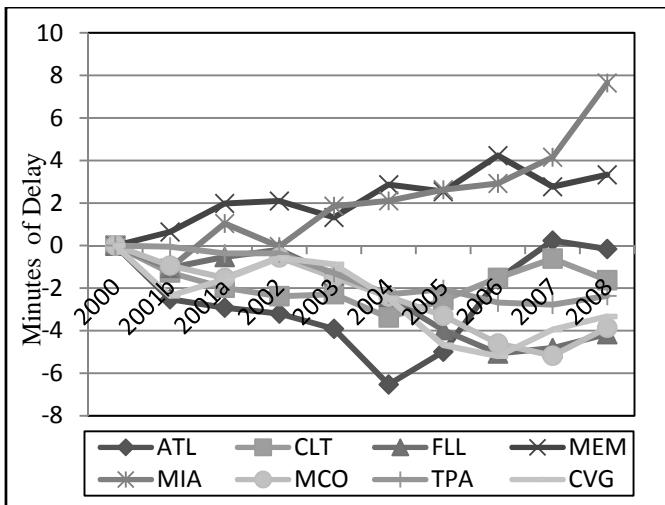


Figure 4. Airport Arrival Delay from 2000-2008 for ASO Region

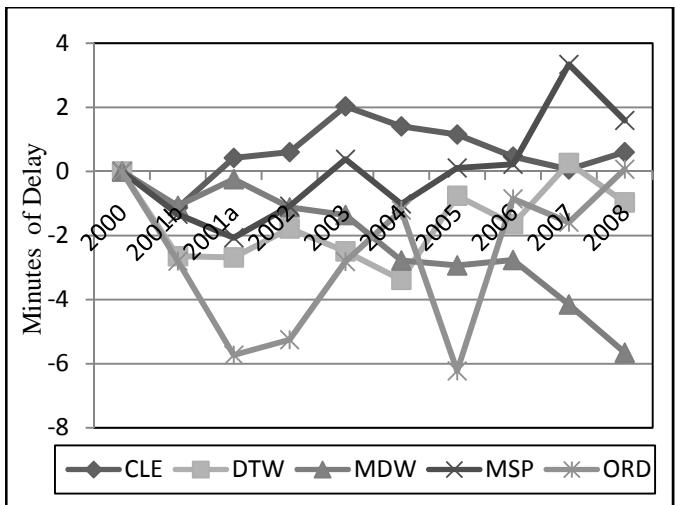


Figure 7. Airport Arrival Delay from 2000-2008 for AGL Region

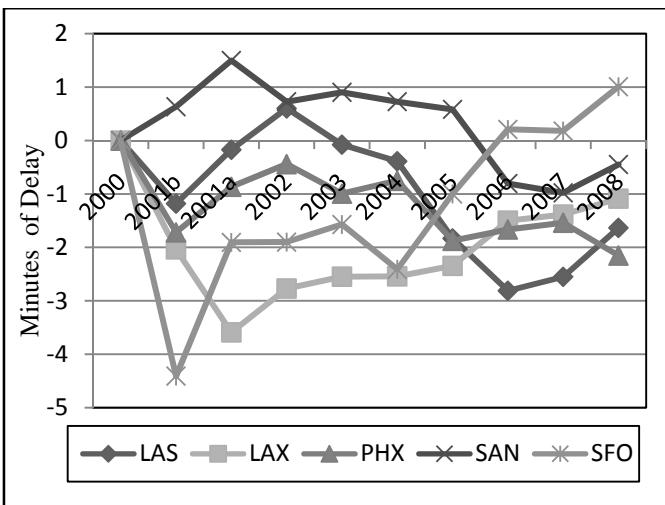


Figure 5. Airport Arrival Delay from 2000-2008 for AWP Region

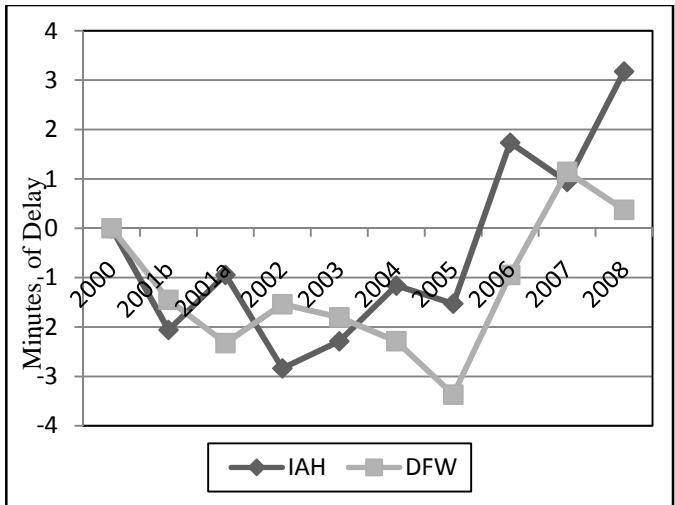


Figure 8. Airport Arrival Delay from 2000-2008 for ASW Region

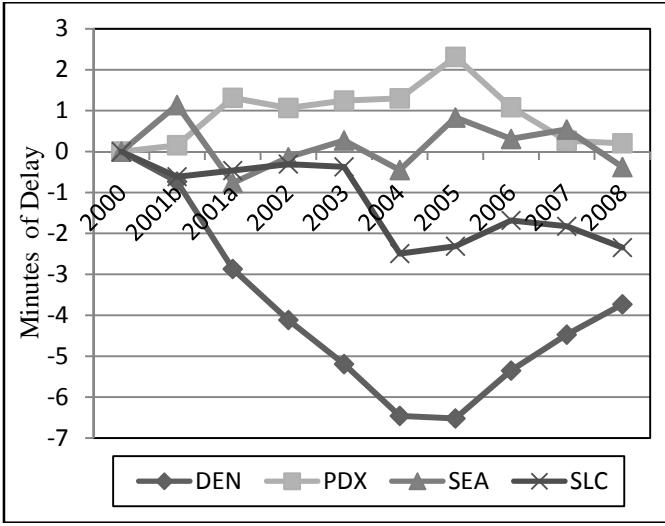


Figure 6. Airport Arrival Delay from 2000-2008 for ANM Region

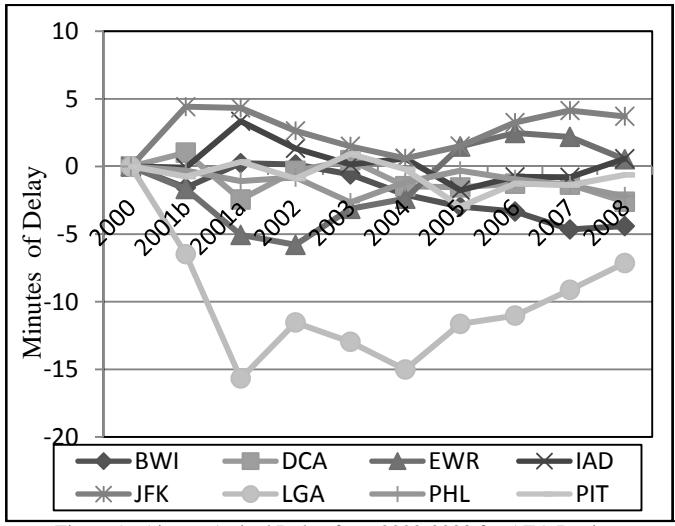


Figure 9. Airport Arrival Delay from 2000-2008 for AEA Region

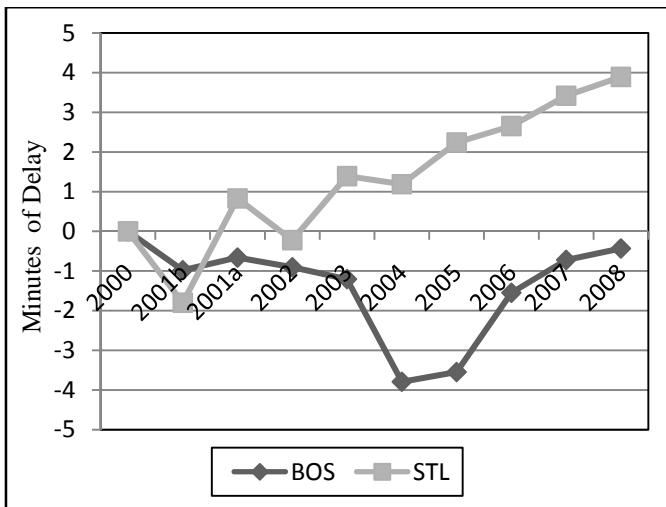


Figure 10. Airport Arrival Delay from 2000-2008 for ANE (BOS) and AAL (STL) Regions

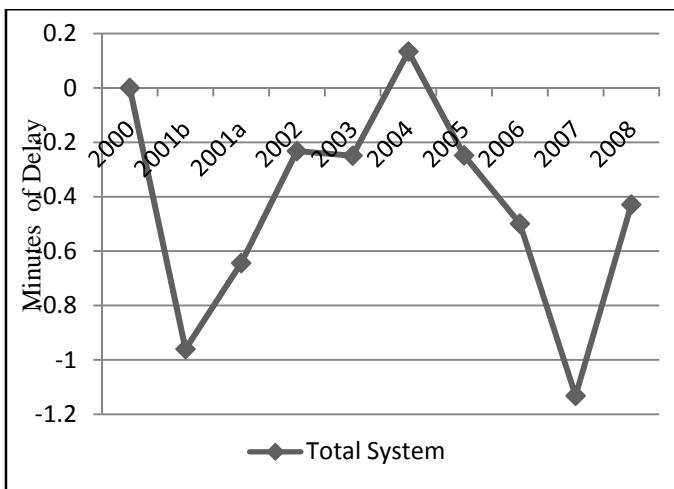


Figure 11. Airport Arrival Delay from 2000-2008 for NAS

LAS	Las Vegas McCarran International
LAX	Los Angeles International
LGA	New York LaGuardia
MCO	Orlando International
MDW	Chicago Midway
MEM	Memphis International
MIA	Miami International
MSP	Minneapolis-St Paul International
ORD	Chicago O'Hare International
PDX	Portland International
PHL	Philadelphia International
PHX	Phoenix Sky Harbor International
PIT	Greater Pittsburgh International
SAN	San Diego International Lindbergh
SEA	Seattle -Tacoma International
SFO	San Francisco International
SLC	Salt Lake City International
TPA	Tampa International

REFERENCES

- [1] Bureau of Transportation Statistics: Airline On-Time Statistics and Delay Causes http://www.transstats.bts.gov/OT_Delay/OT_DelayCause1.asp?pn=1 Accessed: December 15, 2009
- [2] FAA's NextGen Implementation Plan 2009, NextGen Integration and Implementation Office, Washington, DC, 2009 <http://www.faa.gov/nextgen>. Accessed July 10, 2009
- [3] FY CIP 2010-2014 CIP, Estimated Expenditures by BLI (Appendix C), May 2009 http://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/operations/sysengsaf/cip/. Accessed July 4, 2009.
- [4] Beatty, R., Hsu, R., Berry, L., and J. Rome. Preliminary Evaluation of Flight Delay Propagation Through an Airline Schedule, Presented at 2nd USA/Europe Air Traffic Management R&D Seminar, Orlando, FL, December 1998.
- [5] Schaefer, L., and D. Millner. Flight Delay Propagation Analysis with the Detailed Policy Assessment Tool, Presented at the Proceedings of 2001 IEEE Systems, Man and Cybernetics Conference, Tucson, AZ, October 2001
- [6] Schaefer, L., Wang, P.T.R., and Wojcik, L.A. *Flight Connections and their Impacts on Delay Propagation*, Digital Avionics Systems Conference, Vol. 1, October 2003, pp 5.B.4 – 5.1-9.
- [7] Ahmad Beygi, S., Cohn, A., Guan, Y., and P. Belobaba. Analysis of the Potential for Delay Propagation in Passenger Aviation Flight Networks, *Journal of Air Transport Management*, Vol. 14, Issue 5, September 2008, pp. 221-236.
- [8] Zhang, Y., and Nayak, N. A Macroscopic Tool for Measuring Delay Performance in the National Airspace System, Presented at the Transportation Research Board (TRB) 89th Annual Meeting, Washington, DC, 2010.
- [9] Diana, T., Do market-concentrated airports propagate more delays than less concentrated ones? A case study of selected U.S. airports, *Journal of Air Transportation Management*, Vol 15, Issue 6, November 2009, pp 280-286.
- [10] Laskey, K.B., Xu, N., and Chen, C.H. Propagation of Delays in the National Airspace System, Presented at 22nd Conference on Uncertainty in Artificial Intelligence, Cambridge, MA, July 2006.
- [11] Hansen, M., and Zhang, Y. Operational Consequences of Alternative Airport Demand Management Policies – Case of LaGuardia Airport, New York, *Transportation Research Record: Journal of the Transportation Research Board*, No. 1915, Transportation Research Board of the National Academies, Washington, DC, 2005, pp. 95-104.
- [12] Operational Evolution Partnership Topics, Federal Aviation Administration, July 9, 2007 http://www.faa.gov/about/office_org/headquarters_offices/ato/publications/oep/faq/Airports/index.cfm. Accessed January 24, 2010.

APPENDIX

List of Abbreviations related to different OEP airports

ATL	Atlanta Hartsfield International
BOS	Boston Logan International
BWI	Baltimore-Washington International
CLE	Cleveland-Hopkins International
CLT	Charlotte/Douglas International
CVG	Cincinnati-Northern Kentucky
DCA	Ronald Reagan National
DEN	Denver International
DFW	Dallas-Fort Worth International
DTW	Detroit Metro Wayne County
EWR	Newark International
FLL	Fort Lauderdale-Hollywood International
IAD	Washington Dulles International
IAH	George Bush Intercontinental
JFK	New York John F. Kennedy International
HNL	Honolulu International
STL	Lambert St. Louis International

Track 6

Environmental and Weather

Throughput/Complexity Tradeoffs for Routing Traffic in the Presence of Dynamic Weather

Jimmy Krozel*, Joseph S. B. Mitchell†, Anne Pääkkö‡ and Valentin Polishchuk‡§

*Metron Aviation, Inc., Dulles, VA 20166, USA

Email: krozel@metronaviation.com

†Applied Mathematics & Statistics, Stony Brook University, NY 11794, USA

Email: Joseph.Mitchell@stonybrook.edu

‡Computer Science, University of Helsinki, FI-00014, Finland

Email: apaakko@cs.helsinki.fi

§Helsinki Institute for Information Technology, FI-00014, Finland

Email: polishch@cs.helsinki.fi (corresponding author)

Abstract—We present efficient algorithms for computing trajectories for routing multiple aircraft avoiding a set of static or dynamic obstacles (e.g., hazardous weather cells). We present results of an implementation of our algorithms, comparing the throughput and traffic complexity across three routing paradigms:

Static Airlanes: A set of lanes for air traffic is established. The aircraft move in trail along each lane, forming a highly structured traffic pattern. The drawback is that the lanes, being static, may not stay clear of hazardous weather as the weather cells move and potentially block lanes.

FreeFlight: Each aircraft determines its own trajectory in space-time, avoiding moving weather cells and other aircraft. This strategy can result in highly complex traffic patterns that are not amenable to human controller oversight.

Flexible Flow Corridors: This model combines advantages of the static airlanes and the FreeFlight solution. The aircraft are routed along a set of lanes that *slowly change* as the weather cells move. This results in a structured traffic flow amidst moving weather.

Our routing algorithms employ searching in discretized space-time, using a hexagonal packing of disks in free space and a uniform discretization of time. The algorithms allow us to take into account additional routing constraints relevant for Air Traffic Management (ATM).

NOMENCLATURE

ATC	Air Traffic Controller
ATM	Air Traffic Management
BFS	Breadth-First Search
FBRP	Flow-Based Route Planner
FCA	Flow Constrained Area
FEA	Flow Evaluation Area
MIT	Miles-In-Trail
PAZ	Protected Airspace Zone
PC	Popcorn Convection
RNP	Required Navigation Performance
SESAR	Single European Sky ATM Research
SL	Squall Line
SUA	Special Use Airspace
WSI	Weather Severity Index

I. INTRODUCTION

A fundamental task in Air Traffic Management (ATM) is to plan a large number of flight trajectories through a weather-impacted airspace. The capacity of the airspace for a given period of time (planning horizon) is defined as the maximum number of aircraft that can be routed through the airspace during the period, while avoiding hazardous weather cells and respecting the separation standards. This paper investigates the capacity of the airspace and the associated traffic complexity under three routing paradigms: Static Airlanes, Free Flight, and Flexible Flow Corridors. When humans oversee ATM traffic flows, complexity as well as weather hazards will together be the limiting factor in throughput, motivating our study to focus on both throughput and complexity as a function of the amount of weather constraints.

Static Airlanes

The simplest solution to the routing problem is to establish a set of airlanes along which the aircraft travel while respecting the miles-in-trail (MIT) requirements (Fig. 1, left).

Drawback – lanes are static: The airlanes solution is good only in the case of *static* obstacles, as is the case, e.g., with static Special Use Airspace (SUA) constraints. At the same time, the major impact on the capacity of the airspace comes from the weather [1], [2]. Convective weather cells are not static, and as they move they may intersect (and hence – make infeasible) some of the airlanes. Thus, the static airlanes can only serve as a solution on a clear day or during certain periods of time when weather is not overlapping the route structure.

FreeFlight

The aircraft do not have to use a predefined set of airlanes; instead, each aircraft is cleared to fly on its own trajectory (Fig. 1, middle), usually selected in order to take advantage of winds and optimize fuel consumption. The trajectories never intersect among themselves or with the obstacles. Because the planning is done in space-time, the trajectories are designed to avoid *moving* obstacles (provided the obstacles' motion is well predicted from the weather forecast).

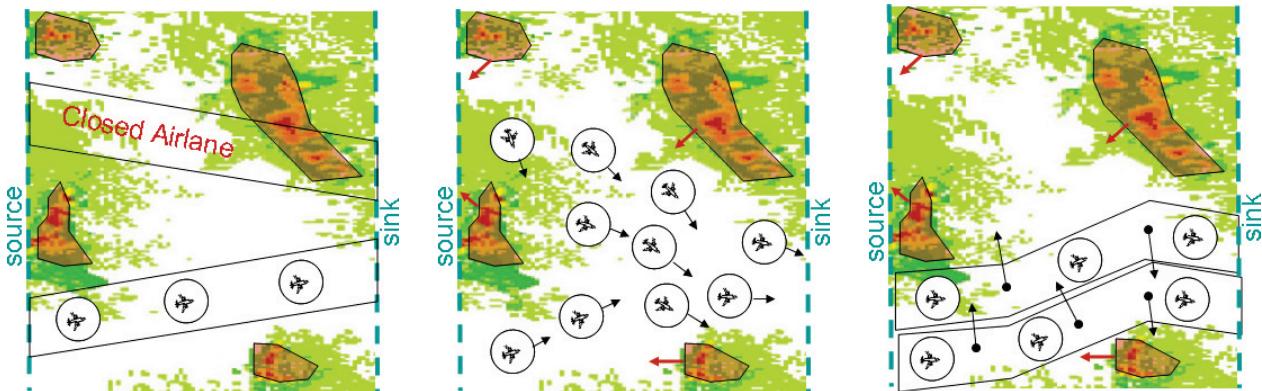


Fig. 1. The rectangle is an airspace; the weather cells are the obstacles for the traffic. Left: Aircraft follow airlanes respecting the miles-in-trail (MIT) standard (5–15nmi). Middle: In the FreeFlight paradigm each jet follows its own trajectory. Right: Flexible Flow Corridors are airlanes that change as the weather cells move.

Drawback – no well-defined airlanes: Prescribing an individual trajectory for each aircraft is far beyond the currently available systems functionality, which are not yet ready for full automation. This is due, in particular, to the presence of humans-in-the loop – air traffic controllers are generally not able to control more than 12–15 trajectories simultaneously crossing one controller’s airspace (sector). Thus, a FreeFlight solution, in which the trajectories of different aircraft can in principle be totally uncorrelated, may turn into an “air traffic controller (ATC) nightmare” – a complicated set of paths tangled in space-time. A controller would prefer the majority of aircraft to follow each other in an orderly fashion, in trail, along a set of (more or less) stable routes – much like it happens in the static-airlanes solution.

Flexible Flow Corridors

Flexible Flow Corridors combine advantages of structured airlanes and of FreeFlight: they feature organized flows and low-complexity traffic amidst moving obstacles. The corridors are thick paths that morph slowly as the obstacles move (Fig. 1, right). Each flow corridor, while morphing, maintains its “threading” through the obstacles. That is, the obstacles that are below the path never “jump” above it, and vice versa. This allows a controller to issue pilot instructions like “Stay north of obstacle 1 but south of obstacle 2”; the instructions remain valid for the entire planning horizon time period.

Motivation: One of the themes of SESAR’s [3] long-term and innovative research (Work Package E [4]) is to explore the possibilities of shifting towards full automation in ATM. In line with this task, we estimate the capacity *ignoring* the possible presence of human-in-the-loop (which corresponds to the full Automation Level 10 [5]). On the other hand, a central role for the human, widely supported by advanced tools to work safely and without undue pressure, is stated as a key feature of the SESAR operational concept. Caring about human factors by bounding the complexity of traffic which a controller will have to monitor, served as our motivation for studying the flexible flow corridors.

A. Related work

Capacity estimation was traditionally done via empirical analysis of controllers’ practices [6], [7], [8], [9], [10]. When weather hazards interfere with static routes, for instance, route blockage or route availability may be determined by analyzing the pilot deviations that are allowed by ATC relative to convective weather cells that overlap jet routes [11]. That is, the capacity estimation was *human-centered*. In contrast, we analyze the capacity *independently* of workload considerations and existing jet routes. We first compute the flow rates *ignoring the controllers’ workload*, and only then measure the complexity of the obtained traffic patterns to see if the complexity is within the controller’s workload limits.

Other related work includes [12], where machine learning techniques were applied to identify routes that survive given the inaccuracy of weather forecasts. Sohier, Bui, and Duong [13] used a packing similar to ours. A grid-based approach to en route dynamic weather avoidance was studied in [14], [15].

B. Our contributions

This paper is a continuation and extension of our prior research on airspace capacity estimation [16], [17], [18], [19]. The novelty of our investigation in comparison with the prior work is four-fold:

- We consider dynamic weather hazards; in [16] the weather hazards were static.
- For static airlanes, in earlier work only algorithms for computing the maximum *number* of airlanes were implemented; here, we report on algorithms that also produce the lanes themselves. Moreover, we find a set of *shortest* lanes. In addition, we produce a set of “conforming” routes, taking into account the geometry of the airspace.
- To find the paths for FreeFlight, previously, the Flow-Based Route Planner (FBRP) [15] was used; here, we implement the algorithm of [19] which allows us to compute trajectories in 3D space-time (i.e., (x, y, t)) whose projections onto 2D space ((x, y)) may intersect each other. (Similarly, FBRP computes paths in space-time,

but does so iteratively, potentially trying different insertion orders; however, it does not have the same theoretical guarantees as the algorithm of [19] that we implemented for our study.) In addition, we show how to use our FreeFlight algorithm to execute holding/airborne delay.

- We study Flexible Flow Corridors; in [16] only static airlanes and FreeFlight were considered.

A minor technical difference is that we use a triangular grid (as opposed to square grid) to discretize the domain.

C. Paper outline

The next section discusses the preliminaries. In Sections III, IV and V we report on the experiments with the static airlanes, FreeFlight, and flexible flow corridors respectively. Section VI compares the throughput and complexity under the three routing paradigms. In Section VII we present extensions of our algorithms, addressing further specifics of motion planning for ATM: we show how to compute routes conforming to sector geometry, and how to use our algorithms to execute holding.

II. MODELING

A. Airspace boundaries

Our focus is on en-route airspace at a constant flight level. In the basic setting, the airspace is modeled as a 300nmi-by-210nmi rectangle; later, we also consider general-shaped airspace representing a sector, center, Flow Evaluation Area (FEA), or Flow Constrained Area (FCA). The traffic enters the airspace through the West (left) side of the rectangle and must exit through the East (right) side. This assumption of mostly unidirectional traffic is justified by the “Alternating Altitude Rule” according to which the flow with West-to-East heading is altitude-separated from the East-to-West traffic.

B. Airspace constraints

The term **obstacle** is used for any region through which flying is not permitted; we do not differentiate between the obstacles induced by the no-fly zones and the constraints induced by hazardous weather cells (possibly, with added safety margins). The Weather Severity Index (WSI) of an airspace is defined as the fraction of its area that is occupied by the obstacles. We experiment with two types of obstacles’ organization: Popcorn Convection (PC) and Squall Line (SL). In the first one (PC), thunderstorms form on a scattered basis, e.g., in the afternoon in response to diurnal heating. The second type (SL) is a solid or nearly solid line or band of active thunderstorms.

To create a computer simulation instance of the routing problem, we populate the airspace with obstacles until reaching the desired WSI. The obstacles are generated using a common distribution across all severity levels; this way the structure of the obstacles can be expected to be similar even as the severity is varied. Each obstacle is a random polygon with 4 to 6 vertices generated in a 40nmi-by-40nmi square (Fig. 2). After the obstacle has been generated, we place it uniformly at random in the airspace: for PC, the obstacle is placed anywhere within the airspace, for SL – only within a vertical band of

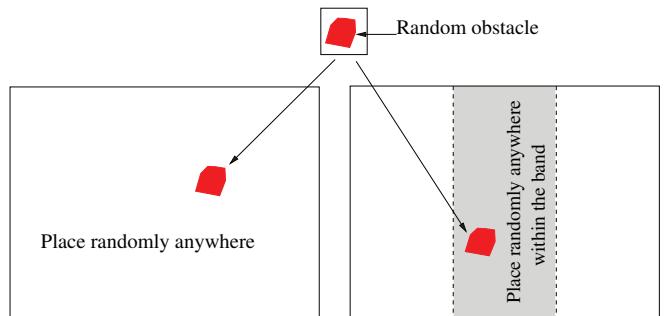


Fig. 2. Obstacles are generated and put into the airspace until reaching the desired WSI. Left: PC. Right: SL.

width 66nmi in the middle of the airspace. Finally, for each obstacle we choose uniformly at random the direction of motion and the speed of motion (this results in a more complicated obstacles motion than is expected in a real-world scenario where the direction of motion of different obstacles are correlated due, e.g., to the wind).

Overall, we experimented with 100 instances for each WSI; for PC we used WSI = 10%, 20%, 30%, 40%, 50%, 60%, 70%, for SL — WSI = 5%, 10%, 15%, 20%, 25%, 30%, 35% (larger WSIs are not possible for SL in our model).

Although we used synthetic weather constraints in the experiments, our algorithms are applicable to real weather data as well. Given a snapshot of real weather, the user may lasso the regions that must be treated as obstacles, or threshold the weather forecast data at a user-specified level based on the observed severity of the weather cells. This is in line with current practices for some airlines, where dispatchers may draw their own boundaries around regions of hazardous weather after viewing convective weather forecasts or turbulence forecasts. The reason for such practices is that there is no straightforward objective criteria to understand which weather cells serve as obstacles for the traffic. In fact, determining the boundaries of hazardous weather is an active research area in ATM; deciding the areas that should be avoided by aircraft is influenced by a multitude of factors – weather cells shape and structure, accuracy of prediction, pilot preference and experience, airline policies, the altitude of radar return echo tops in severe storms, etc. [20], [21], [22].

C. Aircraft trajectories

In our experiments all aircraft travel at a constant speed of 420 kn. The Required Navigation Performance (RNP) requirement for every aircraft is 5nmi (that is, each aircraft can deviate by 5nmi from the route centerline). We assume 10nmi MIT separation. This means that at peak throughput a single airlane can carry 42 aircraft/h past any point in space. Naturally, packing the aircraft “head to tail” is not practical nor realistic. We use this tight packing because our focus is on determining the maximum theoretically possible throughput rates and associated traffic complexity.

Assuming RNP = 5nmi, MIT = 10nmi allows us to model each aircraft as a disk of 10nmi in diameter; the disk represents

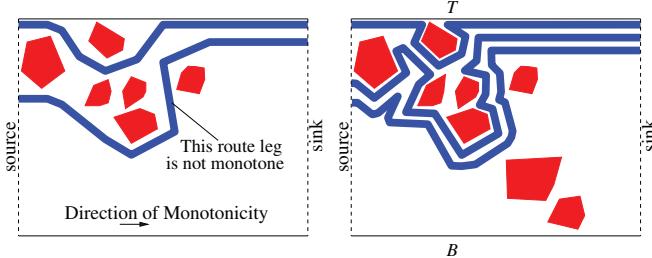


Fig. 3. Left: The northern airplane is monotone (west-to-east), while the southern air lane is not monotone. Right: Uppermost paths fill the space tightly packed starting from the top T to bottom B .

the protected airspace zone (PAZ) around the aircraft. During the motion, the disks have to stay disjoint from each other and the obstacles. Note that we allow the disks to come arbitrarily close to the obstacles (but not to penetrate them); this is because we assume that the hazardous weather regions have safety margins added to them by the user.

It is natural to require that the flight trajectories be monotone in the direction from origin to destination (Fig. 3, left). A flight trajectory will not typically head in the reverse direction except for departure/arrival maneuvers in terminal airspace or flying in a holding pattern.

D. Traffic complexity

We use a variant of the complexity measure from [16] which, in turn, is based on dynamic density investigations [23], [24], [24], [25], [26], [27], [28]. The airspace is tiled regularly with 25nmi-by-25nmi squares, and the time is discretized into 1-minute intervals. For a square p and time t , denote by $A(p, t)$ the set of aircraft that are inside p at time t ; let $|A(p, t)|$ denote the number of the aircraft in the set. The traffic complexity is averaged over all times $t = 1, 2, \dots, 30$. The complexity at time t is the sum of the complexities over all squares. The complexity in the square p is $C(p, t) = 0.36Var(p, t) + 2|A(p, t)|$ where $Var(p, t)$ is the “scaled-contribution” velocity variance, calculated as $Var(p, t) = \sum_{a \in A(p, t)} s_a \|v_a - V_{avg}\|^2$. Here, v_a is the velocity of the aircraft a , $s_a = 1 - |ap_c|/R$ is the scaling factor for a (p_c is the center of p), and

$$V_{avg} = \frac{\sum_{a \in A(p, t)} s_a v_a}{\sum_{a \in A(p, t)} s_a}$$

is the (scaled) “local average velocity”. The final expression for the traffic complexity is

$$\text{complexity} = \frac{1}{30} \sum_{t=1}^{30} \sum_p C(p, t)$$

III. STATIC AIRLANES

The basic routing problem asks for a maximum number of disjoint obstacle-avoiding “thick” paths connecting the source and the sink edges (see Fig. 1, left). Formally a thick path is the Minkowski sum of a usual (thin) path and the disk centered at the origin; the radius of the disk is equal to the RNP. The paths serve as airlanes for the traffic flow.

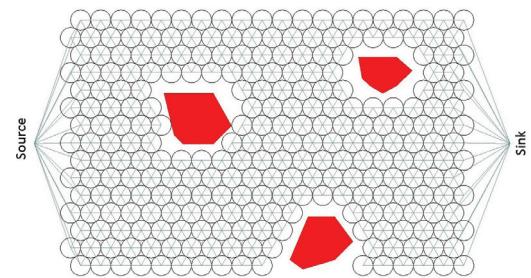


Fig. 4. Domain discretization.

Theoretical solution from prior work: In [19], a continuous-Dijkstra algorithm for computing the maximum number of thick paths in a polygonal domain was suggested. The algorithm runs as follows. The source and the sink edge split the boundary of the outer polygon of the domain into two parts: the top T and the bottom B (Fig. 3, right). The paths are routed in a topmost fashion: the first path runs “as close as possible to T ”, the second – as close to the first as possible, and so on (see Fig. 3, right).

While routing uppermost (or bottommost) paths guarantees that a maximum number of paths will be found, no bound on the *length* of the paths is possible. Another issue is that the algorithm can be hard to implement since the efficient implementation of the continuous Dijkstra method involves waveform tracking, clipping, intersection, etc.

Our solution: We address both of the above issues by discretizing the domain. We start with a hexagonal packing of congruent disks and remove disks intersected by the obstacles. Next, a graph G is formed whose nodes are identified with the disks, and whose edges connect adjacent disks; the graph is a triangular subgrid (Fig. 4).

We connect the disks along the source and the sink to a supersource node and a supersink node and compute a maximum supersource-supersink flow. By the Flow Decomposition Theorem [29], the flow decomposes into a maximum number of disjoint paths in the graph. Moreover, by computing the *minimum-cost* flow, we find a set of *shortest* paths.

Figure 10, left, presents a sample output of our algorithm.

A. Throughput and Complexity

Using the above algorithm, for each problem instance we compute the maximum number of airlanes routable through the airspace at time 0. We then check how many airlanes stay clear of the obstacles during the planning horizon. The throughput is calculated as the number of open airlanes multiplied by the rate of the flow along one lane (42 aircraft/hour) multiplied by the planning horizon length (0.5 hours).

The complexity is calculated as described in Section II-D. Table I presents the results.

IV. FREEFLIGHT

We implemented the algorithm from [19], which finds the maximum number of aircraft trajectories that can be

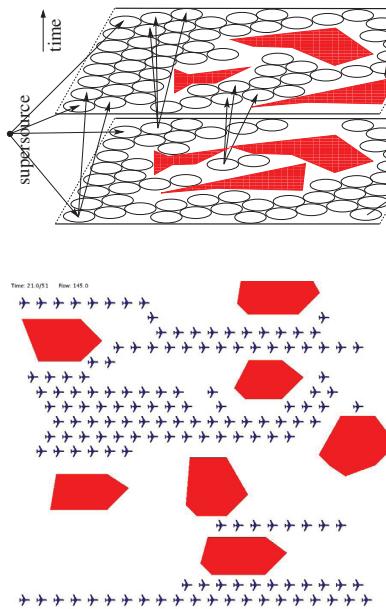


Fig. 5. Top: Motion graph [19, Fig. 13] (not all edges are shown). The time is discretized uniformly, and at every time slice congruent disks are packed in the free space. Each disk is connected to the neighboring disks, on the next time slice, reachable without intersecting the obstacles. Bottom: A snapshot of the implementation output.

routed through the airspace during the planning horizon by computing a maximum flow in the “motion graph” laid out in the (x, y, t) -space (Fig. 5, top). Videos with the output of the implementation can be viewed at our webpage tinyurl.com/ATMexamples¹; Fig. 5, bottom shows a screenshot of the implementation output.

A. Throughput and Complexity

We used the same problem instances as for static airplanes to compute the throughput and complexity for the FreeFlight solution. See Table I for the results.

V. FLEXIBLE FLOW CORRIDORS

We combine the advantages of the static airplanes and the FreeFlight paradigm by considering *Flexible Flow Corridors* – disjoint thick paths that change as the obstacles move, preserving the way they “thread through” the obstacles. Specifically, the *threading* of a path [30] is a vector whose length equals to the number of weather obstacles in the domain, indicating for every obstacle whether the obstacle is above or below the path. We require each flow corridor to stay clear of the moving weather during the planning horizon, and also to keep its threading. We assume that during the planning horizon the obstacles may move and grow/shrink but do not appear or disappear, so the threading is well defined.

Videos with the algorithm’s output can be viewed at tinyurl.com/ATMexamples. Figure 6 compares the FreeFlight solution with the Flexible flow corridors on the same instance.

¹tinyurl.com/ATMexamples is a shorter alias for <http://www.cs.helsinki.fi/group/comgeom/examples.html>

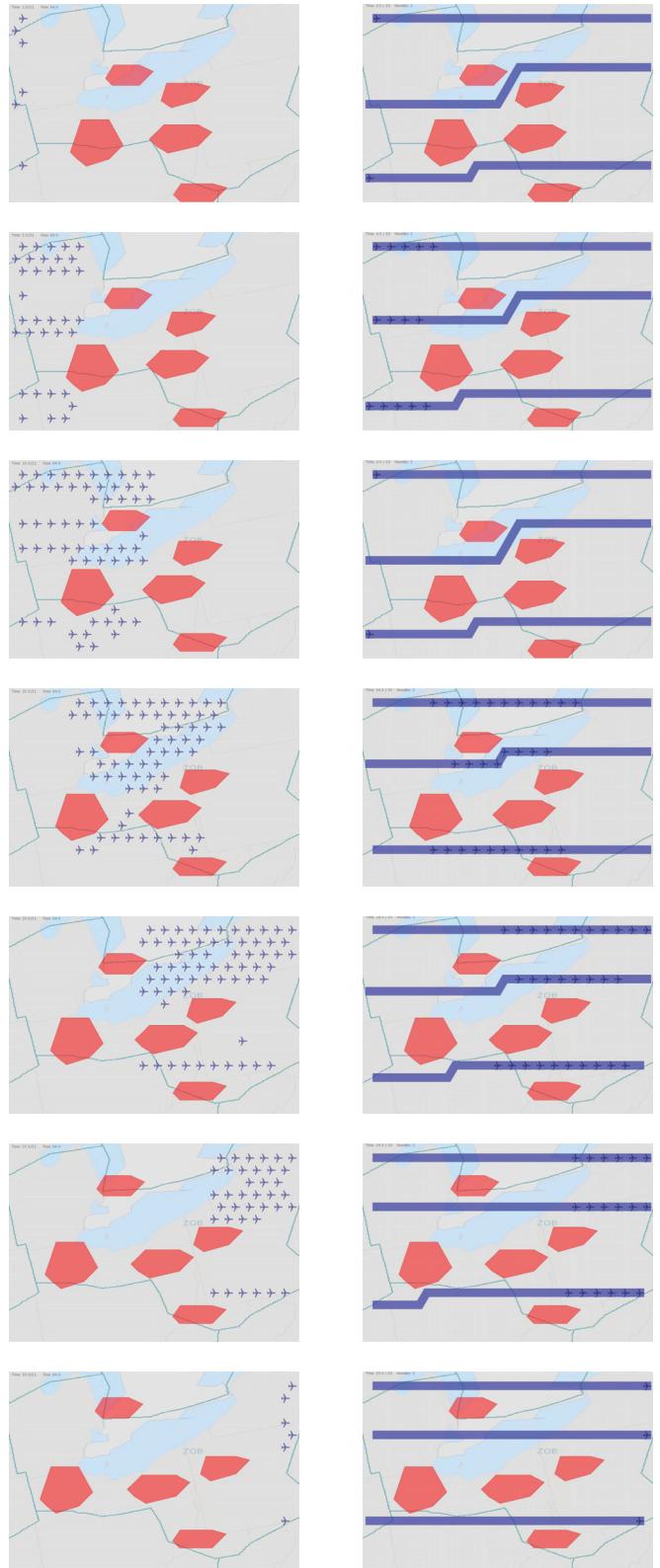


Fig. 6. Snapshots of the airspace at different times (top to bottom): 1min, 5min, 10min, 15min, 20min, 25min, 30min after the start of the experiment). Left: FreeFlight. Right Flexible flow corridors.

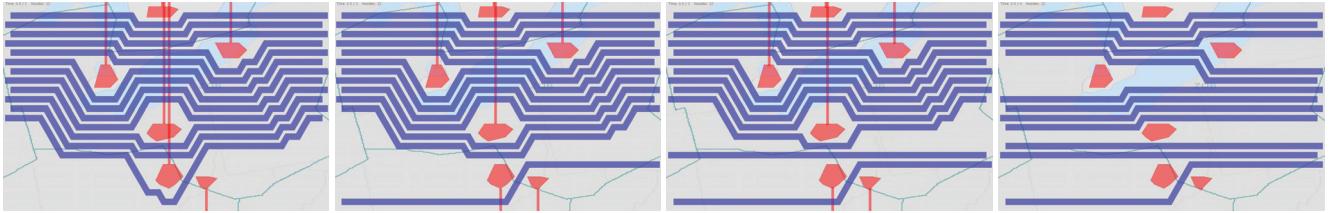


Fig. 7. From left to right: 12 uppermost paths; the obstacles above/below the 12th path are bridged to the top/bottom. The 12th path is shortened, and the obstacles above/below the 11th path are bridged to the top/bottom. 11th path is shortened. All paths are shortened.

Computing the corridors: To compute the corridors we first route uppermost paths at every time slice. As with static airlanes, uppermost paths tend to be unduly long, so we shorten them iteratively and greedily in the “bottommost” fashion: the last (bottommost) path is “pulled taut” treating the next-to-last path as an obstacle, then the next-to-last path is pulled taut treating the path just above it as an obstacle, and so on. Formally, let $\mathcal{P} = (\Pi_1, \dots, \Pi_K)$ be the uppermost paths. Start with the path Π_K , and fix its threading by bridging the obstacles that are above (resp. below) Π_K to T (resp. B); refer to Fig. 7. Now replace Π_K with the shortest path, Π_K^* , in the free space between B and the path Π_{K-1} . (Because the uppermost paths are routed conservatively, in any collection of K paths, the lowest, K th, path cannot intersect Π_{K-1} ; thus, routing Π_K^* as we do seems like a natural idea.) We proceed with the $(K-1)$ st path: after fixing its threading, route the shortest path, Π_{K-1}^* , in the free space between Π_K^* and Π_{K-2} . Continuing this way, we obtain a collection $\mathcal{P}^* = (\Pi_1^*, \dots, \Pi_K^*)$ of shorter paths.

Solution uniqueness: In our implementation, the shortest paths were routed by finding the shortest path in the underlying grid (dual to the disk packing). While in general working with the grid allowed us to overcome many of the difficulties present in continuous versions of the algorithms, shortest paths between two nodes in the grid tend not to be unique. In some instances, this led to the paths “jumping back and forth for no reason” between the neighboring time slices, undermining the very idea of slow morphing.

We thus had to enforce shortest paths uniqueness by perturbing the weights of the grid edges. Easiest to implement was random perturbation; this, however, resulted in the paths looking somewhat random, often making unnatural zigzags. A more consistent perturbation was obtained by introducing a “gravity field”: the edges were heavier towards the middle of the domain thus favoring the flow corridors to run through the middle. This way not only we have enforced the path uniqueness, but also took into account typical controller’s preferences for working with the flights well within the boundaries of their sectors and trying to avoid letting traffic flows get too close to the boundaries.

VI. CONCLUSIONS

Free Flight allowed us to achieve the highest capacity, followed by the flexible flow corridors, followed by the static airlanes (Table I). This is to be expected, since any static airplane that remains open for the planning horizon, is also a

WSI, %	Static airlanes	Free Flight	Flexible corridors	Static airlanes	Free Flight	Flexible corridors
0	165	165	165	220	220	220
10	60.9	93.6	62.5	20 689	38 917	20 877
20	25.8	56.3	27.9	17 154	43 083	18 396
30	14.3	36.9	16.6	8107	29 410	9398
40	5	18.8	7.4	3012	16 217	4638
50	1.5	8.7	2.6	559	7910	1083
60	0.3	4	0.5	270	3414	358
70	0.1	1.1	0.1	0.14	692	0.14
0	165	165	165	220	220	220
5	85.7	110.2	84.6	2384	10 164	2267
10	44.7	74.3	46.6	3305	17 708	2656
15	25.8	55	29	3691	18 433	2537
20	7.7	29.6	9.2	796	12 891	807
25	7.9	20.1	8.3	507	6909	364
30	2.9	10.6	3.1	253	4776	146
35	0.5	3.5	0.6	57	1544	17

TABLE I
THROUGHPUT (LEFT) AND COMPLEXITY (RIGHT) FOR PC (TOP) AND SL (BOTTOM).

feasible corridor (which actually does not morph at all). In turn, a trajectory along a flexible flow corridor is also a feasible FreeFlight trajectory.

As expected, the traffic complexity shows the trend opposite to the capacity (Table I): FreeFlight has highest complexity, followed by the flexible corridors, followed by the static airlanes.

In Fig. 8 the throughput is plotted against the capacity at different WSIs. Overall, we conclude that the throughput and complexity of traffic following the flexible flow corridors is only slightly higher than that of the static airlanes; the throughput and complexity for the FreeFlight is in general considerably higher.

As a tradeoff between throughput and complexity clearly exists, researchers, policy makers and operations personnel (controllers and pilots) need to weigh the compromises between maximizing capacity (throughput) and the potential for complexity-driven safety mishaps. Our research helps to quantify the tradeoffs, and allows one to make the decision on when the complexity is the limiting factor, as well as to estimate the consequences of limiting the capacity due to the complexity being too high.

While our study here has been the off-line evaluation of an airspace to determine tradeoffs between throughput and complexity, we envision that our algorithms can be utilized in evaluation and decision support tools for both off-line and

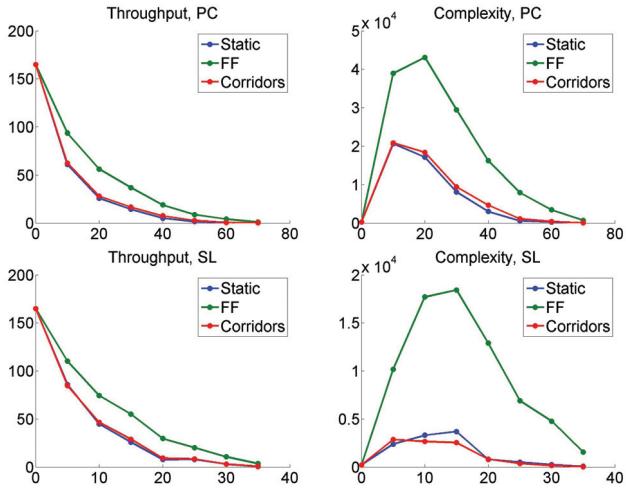


Fig. 8. Throughput (aircraft per 0.5 hr), complexity (unitless) vs. WSI (%).

on-line settings. In an on-line setting, weather forecast data could be analyzed to determine on-line throughput estimates using flexible flow corridors. While the actual computed routes would not be the flown routes, their existence, computed by our algorithms in real time, would provide guarantees of safe throughput, giving the traffic flow decision-maker confidence that, say, 50% (or some other fraction) of the maximum capacity can be safely handled. Several practical considerations will be needed to make this a reality, including the modeling of more operational constraints; further, depending at what level (national, center, sector, etc.) the evaluations are to be done, there may be issues of scale to address.

VII. EXTENSIONS

We emphasize that our algorithmic solutions are applicable to airspaces of various shapes and purposes, without the simplifying assumptions inherent to our basic model. The airspace does not have to be rectangular, but may have an arbitrary geometry. In particular, in Section VII-A below, we consider the case when the airspace represents several adjacent sectors, each with a complicated geometry. We care about the fact that crossing the sector boundary involves a “hand-off” between the sectors controllers. To reduce the communication overhead, we plan flight paths that minimize the number of sector boundary crossings.

Treating the airspace as a 2D (x, y) region makes computing permanent airplanes amidst static obstacles a two-dimensional problem. Taking into account obstacles’ motion translates the problem into 3D (x, y, t). Our dynamic motion planning algorithms for this 3D problem (Section IV) can be extended to plan routes in the full 4D (x, y, z, t) as well; naturally, in this case we will have to find ways of coping with the growth of the number of grid points. In addition, the algorithms work in the case when the obstacles do not just move retaining their shape, but also grow and shrink.

Entering/exiting the airspace only through its West/East side is relevant for en-route airspaces where traffic with different

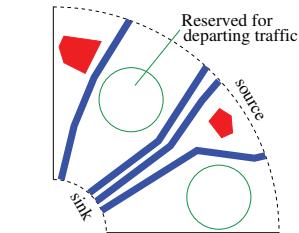


Fig. 9. A transition airspace.

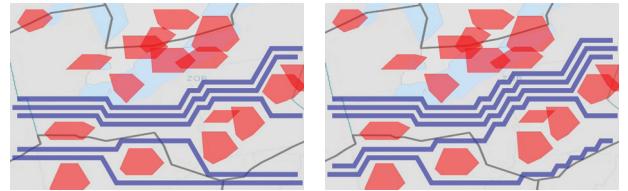


Fig. 10. A sample output on an instance with PC, WSI=35% (the bold line is the sector boundary). Left: 5 paths of minimum total length, computed by the mincost flow. Right: 5 conforming paths minimize the total number of the sector boundary crossings.

headings occupies different flight levels. We can also assume general source/sink areas, including those *inside* the domain. In particular, in Section VII-B below, we consider an instance where the aircraft are initially placed in the airspace *interior* (and have to exit through the East).

Another possible application domain for our algorithms is terminal area operations. The transition airspace can be modeled by a portion of an annulus corresponding to one arrival metering fix (Fig. 9). The source/sink edges are then arcs of the outer/inner range rings – typically of radius about 150nmi/40nmi. While the vertical dimension is crucial here, the 2D (x, y) projection of a full 3D (x, y, z) trajectory may serve as a sufficient specification of the path when the plane follows a predefined descent profile or the altitude is specified along the way-points of the path [31], [32], [33]. In this scenario, in addition to the weather cells and no-fly-zones, the obstacles may represent human-defined “poke-throughs” for ascending/descending traffic following departure routes from an airport or from satellite airports.

A. Conforming flows

The discretization allows us to address additional, ATM-specific constraints. For instance, our airspace may consist of several air sectors, with each sector controlled by a separate controller. Crossing a boundary between adjacent sectors involves a “hand-off” between controllers. Thus it is of interest to find paths that conform to the sectors’ geometry by not crossing the boundaries too often. This is easily achieved in our setting by assigning high costs to those edges of the graph G that cross the boundaries. Figure 10 presents an example.

B. Holding/Airborne delay

Finding a large number of aircraft trajectories, as our FreeFlight implementation does, is already not an easy task

for a controller. It would be even harder for a human to “hold” a set of aircraft within the airspace for some time – such a need arises when the adjacent space downstream the flow is temporarily blocked (e.g., due to equipment failure, a security event, excessive congestion, or extreme growth of weather constraints). Interestingly, the algorithm of [19] allows one to specify the entry and exit time intervals for the aircraft; the disks can appear in and leave the airspace only during the entry and exit intervals. We used this to emulate airborne delay.

The output of our holding pattern implementations can be viewed at tinyurl.com/ATMexamples. In order to make the trajectories more interesting, we made the exit time interval very short; this enforced that before exiting the airspace, all aircraft align along the sink and leave almost simultaneously.

C. Real weather

The output of our implementation on a real-weather instance can be viewed at tinyurl.com/ATMexamples; more extensive experiments with real weather are left for the future work.

D. Demand-driven routing

The methods implemented in this paper focused on routing noncrossing traffic between source/sink pairs on the boundary of an airspace of interest. The traffic demand will, more generally, include traffic that crosses within the airspace and that goes between specified origin-destination pairs. We are generalizing our methods to demand-driven routing.

ACKNOWLEDGMENTS

We thank the anonymous referees for their helpful comments. This research is partially supported by Academy of Finland grant 118653 (ALGODAN). This work is being performed under NASA’s NextGen Airspace Super Dense Operations (ASDO) research focus area under contract NNA07BA84C, from NASA Ames Research Center. We appreciate the financial support of the sponsor of the research, NASA Ames Research Center and the NextGen Project Manager, Dr. Paramal Kopardekar. J. Mitchell is partially supported by the National Science Foundation (CCF-0528209, CCF-0729019).

REFERENCES

- [1] “Your flight has been delayed again,” Joint Economic Committee, United States Senate, Tech. Rep., May 2008.
- [2] Bureau of Transportation Statistics, “Understanding the reporting of causes of flight delays and cancellations,” <http://www.bts.gov/help/aviation/html/understanding.html>, 2008.
- [3] <http://www.sesarju.eu>.
- [4] http://www.sesarju.eu/public/standard_page/wpe.html.
- [5] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, “A model for types and levels of human interaction with automation,” *IEEE Trans. Systems, Man and Cybernetics, Part A*, vol. 30, no. 3, pp. 286–297, 2000.
- [6] D. K. Schmidt, “On modeling ATC work load and sector capacity,” *Journal of Aircraft*, vol. 7, no. 13, pp. 531–537, 1975.
- [7] L. Meyn, “Probabilistic methods for air traffic demand forecasting,” in *AIAA Guidance, Navig., and Control Conf., Monterey, CA*, Aug 2002.
- [8] C. Wanke, L. Song, S. Zobell, D. Greenbaum, and S. Mulgund, “Probabilistic congestion management,” in *6th USA/Europe Seminar on ATM R&D, Baltimore, MD*, June 2005.
- [9] C. Wanke, M. Callaham, D. Greenbaum, and A. Masalonis, “Measuring uncertainty in airspace demand predictions for traffic flow management applications,” in *AIAA Guidance, Navigation, and Control Conf., Austin, TX*, Aug 2003.
- [10] L. Song, C. Wanke, and D. Greenbaum, “Predicting sector capacity for TFM decision support,” in *6th AIAA Technology, Integration, and Operations Conf., Wichita, KS*, Sept 2006.
- [11] B. D. Martin, “Model estimates of traffic reduction in storm impacted en route airspace,” in *AIAA Aviation Technology, Integration, and Operations Forum, Belfast, Ireland*, Sept. 2007.
- [12] D. Michalek and H. Balakrishnan, “Identification of Robust Routes using Convective Weather Forecasts,” in *8th USA/Europe Air Traffic Management R&D Seminar, Napa Valley, CA*, June/July 2009.
- [13] D. Sohier, M. Bui, and V. Duong, “Towards an algorithm to compute safe aircraft trajectories,” in *Proceedings of RIVF*, 2004, pp. 11–14.
- [14] M.-H. Nguyen, S. Alam, J. Tang, and H. Abbass, “Ants-inspired dynamic weather avoidance trajectories in a traffic constrained enroute airspace,” in *6th EUROCONTROL Innovative Research Workshop & Exhibition*, 2007, pp. 205–212.
- [15] J. Prete, “Aircraft routing in the presence of hazardous weather,” Ph.D. dissertation, Stony Brook University, Aug 2007.
- [16] J. Krozel, J. S. B. Mitchell, V. Polishchuk, and J. Prete, “Maximum flow rates for capacity estimation in level flight with convective weather constraints,” *Air Traffic Control Quarterly*, vol. 15, no. 3, pp. 209–238, 2007.
- [17] J. S. B. Mitchell, V. Polishchuk, and J. Krozel, “Airspace throughput analysis considering stochastic weather,” in *AIAA Guidance, Navigation, and Control Conference, Keystone, CO*, Aug 2006.
- [18] J. Krozel, J. S. B. Mitchell, V. Polishchuk, and J. Prete, “Airspace capacity estimation with convective weather constraints,” in *AIAA Guidance, Navigation, and Control Conference, Hilton Head, SC*, Aug 2007.
- [19] E. M. Arkin, J. S. B. Mitchell, and V. Polishchuk, “Maximum thick paths in static and dynamic environments,” *Computational Geometry Theory and Applications*, vol. 43, no. 3, pp. 279–294, 2010.
- [20] R. DeLaura and J. Evans, “An exploratory study of modeling en route pilot convective storm flight deviation behavior,” in *12th American Meteorological Society Conf. on Aviation, Range, and Aerospace Meteorology, Atlanta, GA*, Jan./Feb. 2006.
- [21] B. Martin, J. Evans, and R. DeLaura, “Exploration of a model relating route availability in en route airspace to actual weather coverage parameters,” in *12th American Meteorological Society Conf. on Aviation, Range, and Aerospace Meteorology, Atlanta, GA*, Jan./Feb. 2006.
- [22] W. Chan, M. Rafai, and R. DeLaura, “An approach to verify a model for translating convective weather information to air traffic management impact,” in *AIAA Aviation Technology, Integration, and Operations Forum, Belfast, Ireland*, Sept. 2007.
- [23] “Report of the RTCA board of directors select committee on free flight,” RTCA, Washington, DC, Jan 1995.
- [24] “An Evaluation of Air Traffic Control Complexity, Final Report, NASA Contract No. NAS2-14284,” Wyndemere Corp., Boulder, CO, Oct 1996.
- [25] I. Laudeman, S. Shelden, R. Branstrom, and C. Brasil, “Dynamic density: An air traffic management metric,” Tech. Rep., April 1998.
- [26] B. Sridhar, S. Kapil, and S. Grabbe, “Airspace Complexity and its Application in Air Traffic Management,” in *2nd USA/Europe Air Traffic Management R&D Seminar, Orlando, FL*, Dec 1998.
- [27] R. Mogford, J. Guttman, S. Morrow, and P. Kopardekar, “The complexity construct in air traffic control: A review and synthesis of the literature,” Tech. Rep. Report DOT/FAA/CT-TN95/22, July 1995.
- [28] M. D. Rodgers, R. H. Mogford, and L. S. Mogford, “The relationship of sector characteristics to operational errors,” *Air Traffic Control Quarterly*, vol. 5, no. 4, pp. 241–263, 1997.
- [29] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, 1993.
- [30] J. S. B. Mitchell and V. Polishchuk, “Thick non-crossing paths and minimum-cost flows in polygonal domains,” in *Proceedings 23rd ACM Symposium on Computational Geometry*, 2007, pp. 56–65.
- [31] A. Haraldsdottir, J. Scharl, M. E. Berge, E. G. Schoemig, and M. L. Coats, “Arrival management with required navigation performance and 3d paths,” in *7th USA/Europe Air Traffic Management R&D Seminar, Barcelona, Spain*, July, 2007.
- [32] E. G. Schoemig, J. Armbruster, D. A. Boyle, A. Haraldsdottir, and J. Scharl, “3d path concept and flight management system (FMS) trades,” in *25th Digital Avionics Systems Conf., Portland, OR*, Oct., 2006.
- [33] A. Haraldsdottir, M. E. Berge, L. S. Kang, E. G. Schoemig, M. S. Alcabin, B. W. Repetto, and M. Carter, “Required navigation performance and 3d paths in high-traffic atm operations,” in *25th Digital Avionics Systems Conf., Portland, OR*, Oct., 2006.

Contribution of European Aviation on the Air Quality of the Mediterranean Region: A modeling study

J. Kushta, S. Solomos, G. Kallos
 AMWFG, University of Athens, Greece
 kousta@mg.uoa.gr

Abstract—Aviation is one of the anthropogenic activities with significant past and forecasted growth rate. The emissions from aviation alter the atmospheric composition and have many non neglectable impacts on regional air quality and climate. The study of such processes can be conducted with mathematical models which use an online approach of the meteorological and chemical processes that affect and/or are affected by aviation. The Integrated Community Limited Area Modeling System (ICLAMS) is a fully integrated atmospheric model developed at the Atmospheric Modeling and Weather Forecasting group of the University of Athens. It deals with atmospheric meteorology and chemistry and their interactions in an online coupled way and on the same spatial, temporal and projection platform.

The model was tested for the month of July 2005 for Europe and the Mediterranean Region. Two simulations have been performed, one with emissions from all anthropogenic activities and the second excluding the emissions from aviation. The comparison of the model results, with and without the aviation emissions, gave the opportunity to assess the impact of airport operations on the air pollution levels of the region and downwind areas, under characteristic summer meteorological conditions. The area that is influenced by the emissions from European aviation operations is very large, and the most affected region is the Western and Eastern Mediterranean and several areas in North Africa. The prevailing west – northwest circulation over West and Central Europe favors the transport of pollutants towards East, South East Europe and North Africa leading to perturbations in the atmospheric composition especially up to 4 – 5 km above surface. The ozone field is altered by the aviation emissions with perturbations in its daytime values that reach 5 - 8 ppb. The atmospheric concentrations of other gas and aerosol pollutants are also affected.

Keywords-air quality, pollution models, Athens Airport, emission inventories, aviation emissions

I. INTRODUCTION

Transportation (land transport, shipping, aviation) is an important contributor to regional and global emissions of several greenhouse gases and aerosols which play a crucial role, directly and indirectly, on clouds, radiation and climate (IPCC 2001). Aviation is one of the transport sectors with a remarkable growth rate that is expected to continue growing.

This growth raises concerns on the environmental impact associated with aviation, including air pollution and noise, stratospheric ozone, radiation and cloud impacts, from regional to global spatial scales and from seasonal to decadal timescales. The attention has shifted from stratospheric ozone depletion from supersonic aviation in the 1970s, to upper troposphere – lower stratosphere (UTLS) ozone enhancement from subsonic aviation emissions during the 1990s. Recent research focuses on cloud cover perturbations (aerosol activation as CCN and IN) and radiative forcing (absorption and scatter of solar radiation) of aviation emissions. On the other hand, environmental issues can constrain air transport through adverse meteorology and environmental performance regulations.

The meteorological and chemistry processes that take place in the atmosphere are usually treated mathematically in an offline approach with non-existing or inadequate representation of significant microphysical and radiation processes and the interaction between these processes and the pollutants fields. Under specific conditions, this variability and the feedback of the chemical component on the meteorological processes becomes a dominant factor in the evolution of the air pollution field and leads to alterations of the cloudiness and radiation transfer budget. In this study a new approach for the treatment of the aviation emissions in a regional atmospheric model has been used.

The ICLAMS model is an online coupled fully integrated, meteorology – chemistry – aerosols – cloud radiation atmospheric modelling system (Kallos et al. 2009; Kushta et al., 2008). It comprises of four submodels: a pre-processing, two core meteorological and chemical and a post-processing submodel. The pre-processing unit prepares the emission fields for the domain of interest, from different emission inventories (local, regional, global) converting them to the spatial resolution and the projection of the meteorological and chemical unit (polar stereographic). The meteorological unit is based on the Regional Atmospheric Modeling System (RAMS) (Pielke et al., 1992; Cotton et al. 2003). Its explicit microphysics parameterization provides detailed information

on the cloud properties allowing a more accurate representation of the cloud-aerosol interaction processes. The chemical unit treats the photochemical reactions, the gas, aqueous and aerosol phase chemistry mechanism and the atmospheric cycle of the online emissions of natural aerosols (mineral dust, sea salt) and biogenic emissions (isoprene). The natural aerosols and the chemical aerosols of sulfates and nitrates have been coupled with the radiation budget scheme and the microphysics scheme, accounting for the radiative forcing and CCN activation of these aerosols.

This paper is organized as follows: A description of the model characteristics is given in section II. In section III an evaluation of the new modeling system is performed. Section IV analyzes the case study regarding the contribution of the European aviation emissions on the air pollution field of Europe and Mediterranean. Section V summarizes the main findings.

II. MODEL CHARACTERISTICS

For the development of the new modeling system the Regional Atmospheric Modeling System – RAMS (version 6.0) is used as the base meteorological model, providing the dynamic cores for the advection, diffusion, radiation, clouds and surface processes. The nesting capabilities of RAMS allow the model to solve large domains of low resolution together with high and very high resolution nested ones. The explicit cloud microphysics scheme provides detailed information on the cloud properties adding to the accuracy of the wet deposition process, and aerosol – cloud interaction.

The first part of the new development includes the implementation of the natural and anthropogenic emissions. The atmospheric cycle of natural emissions includes the mobilization, transport and removal processes for desert dust and sea salt. These two categories of natural particles are important to aviation for many reasons. They can affect aviation in a direct way through reduced visibility or in an indirect way through dust ingestion by the jet engines. Another category of natural emissions are the VOC emissions from vegetation. Isoprene is the main chemical species emitted by vegetation and its emission is described with the method proposed by Guenther et al. 1993.

The anthropogenic emissions database is provided by the Joint Research Project (Ispra, Italy) and includes the gas pollutants, such as NO, NO₂, SO₂, CO, NH₃, VOCs and aerosol pollutants OC and BC at a resolution 0.1x0.1 latitude longitude. The aviation emissions include the same pollutants and are given in 4 operation altitudes: Land – Take off cycle 0 – 1 km, Climbing and descend 1 – 9 km ,Cruise 9 – 13 km, Supersonic > 13 km.

The chemical component of the model constitutes of the gas and aqueous chemistry module, the gas-particle interaction module, the transport and removal modules. The gas chemistry module is based on the chemistry mechanism SAPRC99 (Carter et al. 1988, 1990). The photochemical scheme uses the basic formulations proposed by Madronich et al. (1987). They are calculated directly on the detailed 3D temperature and pressure fields from RAMS model at a user defined frequency of 600 seconds. The aqueous chemistry module deals with the removal of the pollutants through scavenging and wet deposition, and with the chemical processes that take place inside a cloud. For the gas-aerosol processes the ISORROPIA mechanism is incorporated into the model. The mechanism used in this development includes ammonium, sodium, chloride, nitrate, sulfate and water, which are partitioned between gas, liquid and solid phases (Nenes et al. 1998).

III. TEST CASE JULY 2005

A test case for July 2005 has been performed, in order to study the ability of the coupled modeling system ICLAMS to capture the temporal and spatial variability of important air pollutants in the Mediterranean – Europe region and the contribution of aviation emissions on their atmospheric distribution. During this period there is high photochemical activity over the region of interest, low cloud cover and high irradiances. The domain of study has been set to have the characteristics given in Table 1. Ozone is the main gas pollutant used in this study because of its importance in defining the air quality index of an area. Sulfate aerosols are also analyzed due to their impact on the cloud properties (by acting as CCN) and the radiation budget of the Earth.

The initial conditions used as the starting point for the model calculation of the concentrations of air pollutants are taken from a lookup table with the value for background ozone being 35 ppbv. A spin off time of 48 hours is given to the model with the analysis period being from 03 till 30 July. An initial statistical performance study has been performed with mean hour surface ozone observations from more than 60 EMEP station around Europe. These are stations with different characteristics, such as different altitudes, geographic location (shore, mountain, valley etc) and background classification (rural, background rural, semi urban etc). The mean hour surface ozone concentrations are compared in all stations. The statistics used for the analysis of the results include the mean bias which states the systematic error for a continuous variable and is defined as:

$$bias = \frac{1}{N} \sum_{i=1}^N (F_i - O_i) = \bar{F} - \bar{O}$$

and the correlation coefficient – R which describes the correlation level between the model and observation values and

is defined as the ratio of the covariance of the values to the product of their standard deviations:

$$r_{F,O} = \frac{\text{cov}(F, O)}{s_F \cdot s_O}$$

The mean bias (units ppbv) factor quantifies the analogy between the mean modeled value and the mean observed value and can take negative or positive values. If bias < 0 then the model underestimates the set of values of the variable while bias > 0 means that the model has an overestimation over the observed mean value. The correlation coefficient can vary from -1 to + 1. The best correlation of the observed versus modeled values of the variable is achieved when the correlation coefficient reaches the value +1.

All the statistical results are summarized in Figure 1 and refer to the time period 03-30 July 2005. The correlation coefficients show good agreement between modelled and observed mean hour surface ozone for the majority of the EMEP stations. The mean biases for the test period are mainly negative for most of the stations. This is an indication of a slight underestimation of the mean hour ozone values that are predicted by the model in this area. There are several factors that can contribute to this outcome. One factor may be the initialization of the model from lookup tables with one value for the whole domain. Another significant factor may be the accuracy of the emission inventories and especially the temporal (daily and hourly) factors that are applied in order to convert the monthly emissions to daily and hourly values.

An additional parameter to better statistical results is usually the spatial and temporal resolution of the model configuration. The large resolution does not help in capturing maximum ozone levels as well as abrupt changes in ozone concentrations which depend on very local geophysical characteristics and localized traffic emissions. Long range transport of pollution from distant areas, which are outside of the model domain, can lead to underestimation of the pollutants field under specific meteorological conditions. Despite all these uncertainties, the comparison shows that the model captures in a satisfactory way the variation of the mean hour concentrations in European stations.

IV. CONTRIBUTION OF EUROPEAN AVIATION ON MEDITERRANEAN POLLUTION FIELD

In order to identify the contribution of the airport emissions on the ozone levels of Europe and Mediterranean another configuration has been set. Using the same domain characteristics and chemical and physical mechanisms as before, the air pollution field of the area has been simulated zeroing out the aviation emissions of BC, CO, NOx, OC and SO2 in the aviation emissions inventory from JRC. With this

configuration the air pollution field of the area of interest has been modified.

The meteorological conditions during the test period present a complex wind field over Europe with an anti-cyclonic circulation over West Europe, during the first days of the test period, and a prevailing westerly flow during the rest of the days (Figure 2). The westerly flow is also present over Mediterranean especially during the second half of the test period. The surface flow pattern favors the transport of anthropogenic pollution produced over Europe and West Mediterranean towards East Mediterranean and North Africa, preventing at the same time the transport of natural aerosol pollutants like mineral dust towards higher latitudes

The alteration in the ozone field over Europe during day time and Mediterranean comes as a result of two mechanisms. Firstly, by excluding aviation emissions from the emission inventory of the region, less ozone precursors are emitted since NOx, CO and VOCs from airports are neglected. This leads to less ozone formation in and around airports in this case. Secondly, by not including aviation emissions, less ozone precursors are transported downwind leading to decreased amounts of ozone formed in distant areas due to less availability of its precursors. During night time the ozone field remains higher when the aviation emissions are included for two reasons: due to the preservation of the daytime difference and due to less reaction between ozone and NO.

Since the ozone formation reactions are fast reactions, the ozone transport is the main factor of the modification of the air pollution field in these downwind areas both during day (Figure 3a) and night (Figure 3b). This explains the linear and one sign alteration (positive = increase of ozone concentrations when the aviation emissions are included in the inventory) of the ozone air pollution field in these distant areas. The ozone difference patterns are dictated by the circulation patterns. For example when the westerlies are enhanced over West Europe the pollution is transported towards East Europe and further down towards Eastern Mediterranean due to the seasonal north winds (trade winds) over Aegean giving higher ozone concentrations over East Mediterranean and North-East Africa. The change of the wind field into an anticyclonic pattern over Central Europe and Italy leads to an enhanced transport of pollution towards Central Mediterranean and North coast of Africa and even inland Africa, before turning back to the Iberian peninsula leading to increases in ozone concentration over Spain and Italy of up to 5 – 6 ppbv. Respectively, the differences in the sulfate aerosols concentrations can reach 0.3 ug/m³ at noon time, in the case where the aviation emissions are included because of the elevated SO2 emissions. Due to limited cloud cover the main mechanism for the production of sulfate aerosols is the oxidation of the emitted SO2 rather than in cloud sulfate formation and heterogeneous mechanisms. Hence the surplus of SO2 emissions from aviation leads to an

increase in sulfate aerosols especially downwind from the airports areas. These results are illustrated in Figure 4 where the time series for mean hour ozone and sulfate aerosols for four main European cities in the vicinity of major airports are shown. In these cities, both ozone and sulfate aerosols in the case when the airport emissions are included in the simulation are higher than the respective concentration when the emissions from airports are not included. The differences in ozone concentrations vary from 5 ug/m³ for Madrid, Athens and Rome, to 10 ug/m³ for Paris Metropolitan area while sulfate aerosol increase reaches 10 – 15 % of the background value. Paris is located southwest of its main airport (Charles de Gaulle) and during this time period the main wind direction is north – northeast maximizing the impact of the inclusion of the airport emissions on the city air pollution burden.

V. SUMMARY AND CONCLUSIONS

A coupled online meteorological and chemical modeling system has been developed for a more accurate assessment of air quality patterns in a local and regional scale. This integrated modeling tool can be used either as operational (high resolution weather and air pollution forecasting), or as scenario and policy tool (environmental performance of aviation) or as a training asset. In this study, an effort has been dedicated to the identification of the emissions from European aviation on the air pollution field over Europe and Mediterranean.

Under characteristics summer conditions emissions from European aviation can influence not only Europe itself, but the Mediterranean Region and North Africa as well, and possibly even far more distant areas. The aviation emissions as produced by JRC can increase the mean hour ozone concentrations during noon time by approximately 2 – 10 ug/m³ with the most affected areas being the Iberian Peninsula, Central and East Mediterranean. Other pollutants, such as sulfate aerosols show an increase due to SO₂ emissions from the airports operations. Downwind areas are more vulnerable to such emissions, and especially when these areas are highly populated and/or industrialized. The perturbations in the concentrations of the ozone precursor such as VOCs, CO and NO_x and particularly the ratio between VOCs and NO_x determine the level of impact of the airport emissions over these areas. Large cities in the vicinity of major airports can show higher ozone differences, which can reach 10 ug/m³ under favorable meteorological conditions.

Due to the complex terrain and topographic characteristics of the area of study, the modeling of the air pollution patterns must be performed with high resolution configuration. This can be realized with the use of nested (smaller incorporated grids into the parent grid). A significant drawback to using high resolution configuration in air pollution models is the computational time required for the solution of the parameterization equations for each grid and the interaction

processes between this grids. Despite this difficulty, the ICLAMS has been developed in order to perform high resolution modeling with parallel processing. In that case, the nested grids with higher resolutions can use other emission inventories which are expected to give a more accurate representation of the air pollution field of the area of interest. Aviation emission scenarios can also be applied in order to assess their impact on air quality in the short and long term.

Future work includes the performance of such assessment tests for winter and transition seasons (spring, autumn) and sensitivity tests for different scenarios on increase/decrease of all or particular gas pollutants emissions from aviation with feedbacks on clouds, precipitation and radiation budget.

ACKNOWLEDGMENTS

This work is supported by EUROCONTROL, under the terms of the Research Studentship Agreement no. C06/22048ST and the EU 6th Framework Program CIRCE IP, contract# 036961.

REFERENCES

- [1] Carter W. P. L. (1988) Documentation of the SAPRC Atmospheric Photochemical Mechanism Preparation and Emissions Processing Programs for Implementation in Airshed Models. Final Report for California Air Resources Board Contract No. A5-122-32
- [2] Carter W. P. L. (1990) A detailed mechanism for the gas-phase atmospheric reactions of organic compounds. *Atmos. Environ.* 24A, 481-515
- [3] Cotton, W. R., R.A. Pielke Sr., R.L. Walko, G.E. Liston, C.J. Tremback, H. Jang, R.L. McAnelly, J.-Y. Harrington, M.E. Nicholls, G.G. Carrio, and J.P. McFadden, (2003) RAMS 2001: Current status and future directions. *Meteorol. Atmos. Phys.*, 82, 5-29
- [4] Guenther, A., P. Zimmerman, P. Harley, R. Monson, R. Fall, Isoprene and monoterpene emission rate variability: Model evaluation and sensitivity analysis, *J. Geophys. Res.*, 98, 12609-12617, 1993.
- [5] IPCC, 2001. Intergovernmental Panel on Climate Change's Synthesis Report. Climate Change 2001 or web versions of 2001 IPCC reports.
- [6] Kallos G., S. Solomos, J. Kushta 2009: Air Quality – Meteorology interaction processes in the ICLAMS modeling system. 30th ITM Proceedings – Conference on Air Pollution Modeling and Applications
- [7] Kushta J., S. Solomos, C. Spyrou, M. Astitha, E. Mavromatidis, G. Kallos, 2008: A modeling system for describing mineral dust and sea salt cycles in the atmosphere. 12th ATRS World Conference, Greece
- [8] Madronich S., 1987, Photodissociation in the atmosphere, 1, actinic flux and the effects of ground reflections and clouds. *Journal of Geophysical Research*, 92, 9740-9752
- [9] Nenes, A., C. Pilinis, S. N. Pandis, 1998. ISORROPIA: A new thermodynamic model for inorganic multicomponent atmospheric aerosols. *Aquatic Geochemistry*, 4, 123-152.
- [10] Pielke, R. A., W. R. Cotton, R. L. Walko, C. J. Tremback, M. E. Nicholls, M. D. Moran, D. A. Wesley, T. J. Lee, J. H. Copeland, 1992. A comprehensive meteorological modeling system – RAMS. *Meteorol. Atmos. Phys.*, 49, 69-91
- [11] Walko R., C. Tremback "RAMS Regional Atmospheric Modeling System: Introduction to RAMS 4.3/4.4". Colorado State University

FIGURES AND TABLES

Table 1. Model configuration

Number of X points	210
Number of Y points	160
Number of levels	29 (with vertical resolution starting from 100 m to 1000 m and model height up to 18 km)
Horizontal resolution	24 km
Temporal resolution	30 sec
Domain center	lat = 39° lon 16°
Chemistry options	Gas, aqueous and aerosol chemistry activated

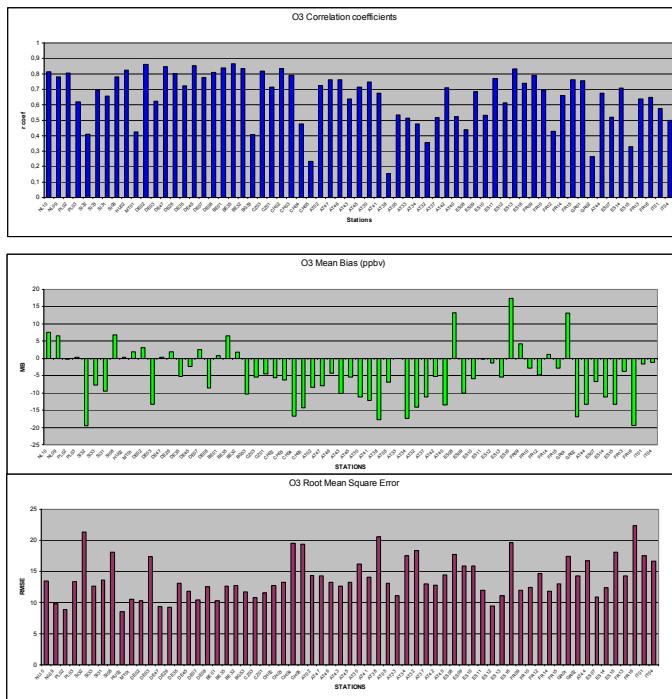


Figure 1. Correlation coefficients, Mean bias and Root Mean Square Error from the comparison of modeled values of mean hour surface ozone concentrations with observations from EMEP stations for the time period 03-30 July 2005.

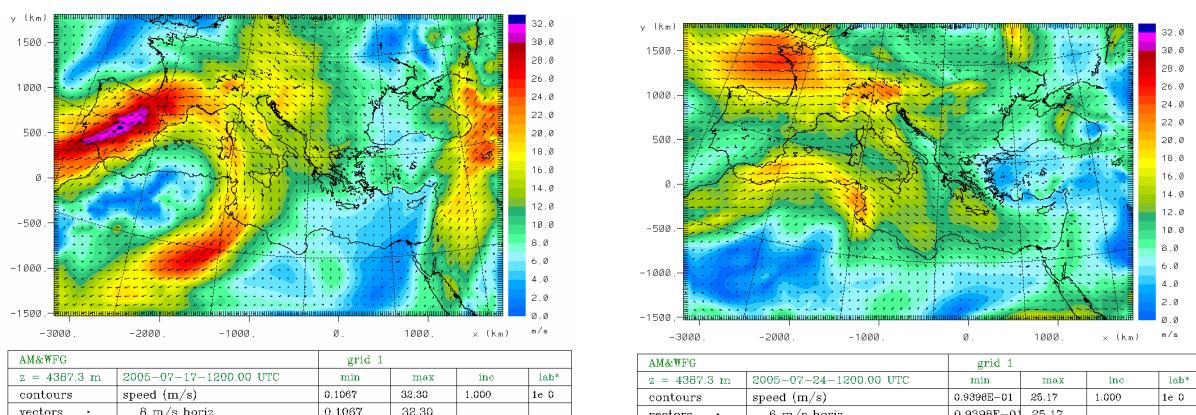


Figure 2. Wind speed and direction at 4500m over Europe and Mediterranean during the test period for 17 and 25 July 2005

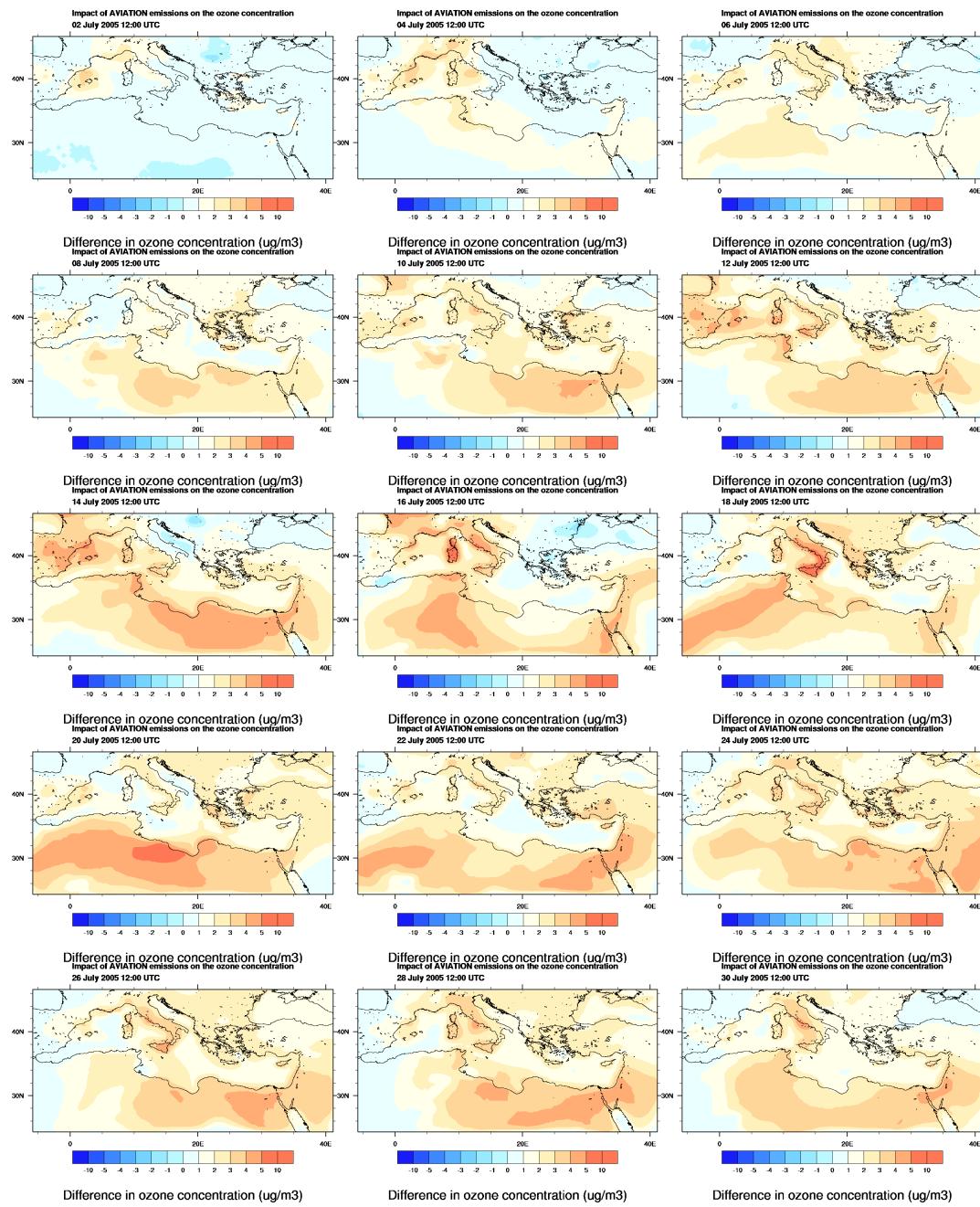


Figure 3a: Difference in the mean hour surface ozone concentration due to the utilization of the aviation emissions in the emission inventory for the even dates of July 2005, at 12:00 UTC. The difference is expressed as Concentration of ozone with aviation emissions minus Concentration of ozone without aviation emissions.

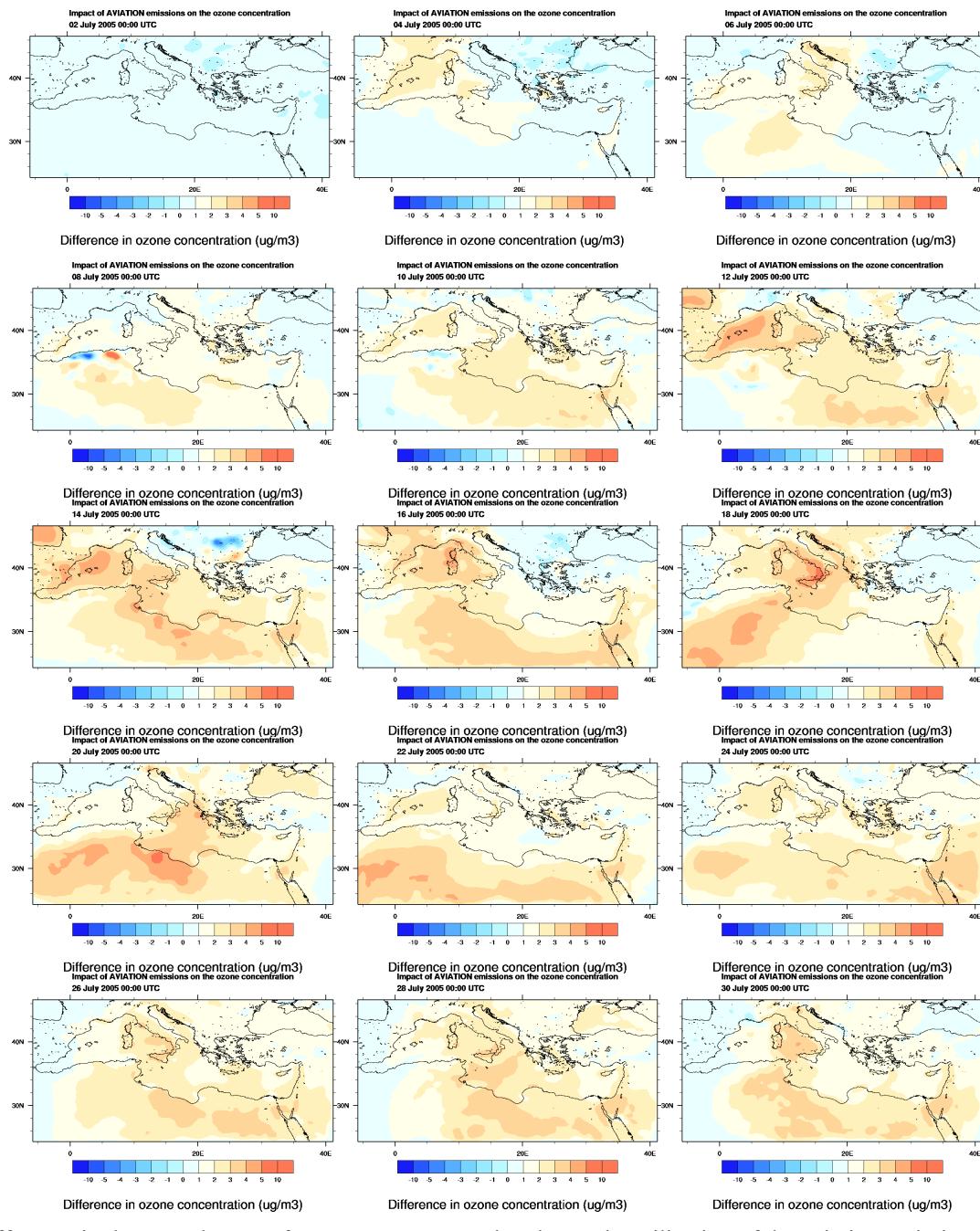


Figure 3b: Difference in the mean hour surface ozone concentration due to the utilization of the aviation emissions in the emission inventory for the even dates of July 2005, at 00:00 UTC. The difference is expressed as Concentration of ozone with aviation emissions minus Concentration of ozone without aviation emissions.

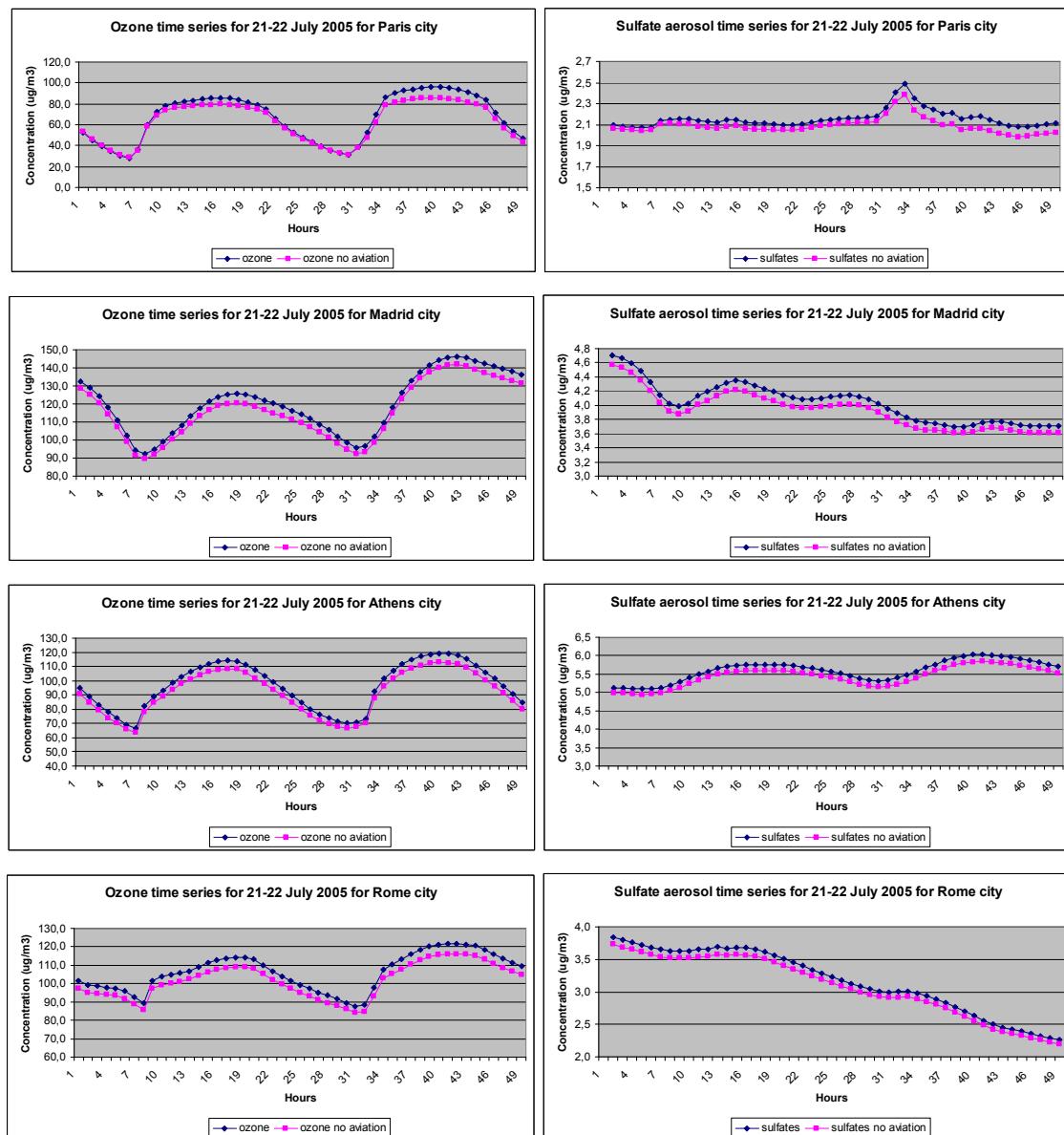


Figure 4. Ozone (left) and sulfate aerosols (right) time series during 21 and 22 July in four major European cities (Paris, Madrid, Athens, Rome) located near airports with significant emissions of gas and aerosol pollutants.

Generating Day-of-Operation Probabilistic Capacity Profiles from Weather Forecasts

Gurkaran Singh Buxi, Mark Hansen

Department of Civil & Environmental Engineering
University of California, Berkeley

Berkeley, California, USA

gkb@berkeley.edu, mhansen@ce.berkeley.edu,

Abstract—It is common understanding that weather plays an important role in determining the capacity of an airport. Severe weather causes capacity reductions, creating a capacity demand imbalance, leading to delays. The role of air traffic flow management (ATFM) measures is to reduce these delay costs by aligning the demand with the capacity. Ground delay program (GDP) is one such measure. Though the GDP is initiated in poor weather conditions, and weather forecasts are subject to errors, present GDP planning procedures are essentially deterministic in nature. Forecast weather is translated into deterministic capacity predictions on which GDP planning is based. Models which employ probabilistic capacity profiles for planning GDPs have been developed, but their application has been limited by the inability to create such profiles from weather forecasts. This paper focuses on San Francisco International Airport (SFO) and provides a methodology to generate probabilistic capacity profiles from the two terminal weather forecasts: Terminal Aerodrome Forecast and San Francisco Marine Initiative (STRATUS). The profiles are inputs to a static stochastic GDP model to simulate ATFM strategies. The solution from the model is evaluated against realized capacities to determine the benefit of the forecast. The benefit of inclusion of the weather forecast is assessed by comparing costs of delays from ATFM strategies simulated from probabilistic profiles developed without the weather forecasts. It is also shown that inclusion of weather forecasts reduces the cost of delays. The paper also compares the cost of delays from strategies simulated using the profiles generated from TAF and STRATUS. It is shown that on average TAF offers similar benefit in controlling cost of delay when compared to STRATUS, indicating that other airports would also benefit from using TAF in planning of operations.

Keywords- Air traffic flow management; Ground delay program; Probabilistic Capacity Profiles; K-means Clustering; Terminal Aerodrome Forecast; STRATUS

I. INTRODUCTION

Adverse weather conditions in the vicinity of an airport reduce the operational capacity of the airport leading to an imbalance between capacity and demand. This capacity-demand imbalance creates en-route airspace congestion leading to delays, an increase in cost and elevating Air Traffic Flow Management (ATFM) risk. If adverse weather is present in the vicinity of the airport, the Federal Aviation Administration (FAA) implements Ground Delay Programs (GDPS). GDPS mitigate weather-induced airspace congestion by metering the

arrival of aircraft to the destination airport. The metering matches the number of flights arriving in a period with the “airport acceptance rate” (AAR) forecast. The metering of flights is achieved by delaying inbound flights on ground at the origin airport prior to their departure. If the AAR forecast is perfectly accurate, the metering from the GDP ensures that the total delay costs are minimized.

It is common understanding that the AAR is primarily influenced by the weather in the vicinity of the airport and thus AAR forecasting necessitates a terminal weather forecast. The weather forecasts are seldom accurate in perfectly predicting the conditions and can thus lead to inaccurate predictions of the AAR. GDP models found in the literature incorporate the uncertainty in the AAR and can be classified in two broad categories: dynamic models and static models. In dynamic models as the information is updated, ground holding decisions are revised, incorporating a wait-and-see strategy. Dynamic models require scenario trees to represent the uncertainty in the AAR. Conversely, in a static model, decisions made once are not revised. Static models require probabilistic capacity profiles as inputs. Reference [1] contains more details on the types of GDP models. A substantial amount of theoretical research on GDP models has been found. However, capacity profiles or scenario trees that the models assume are often used for illustrative purposes and are hence not considered to be a part of ATFM decision making. There exists a gap in literature on the development of specific day-of-operation probabilistic capacity profiles, hindering incorporation of tailor-made day-of-operation strategies in ATFM. This paper focuses on the development of probabilistic capacity profiles generated from day-of-operation weather forecast. The potential benefit of incorporating weather forecasts into day-of-operation would lead to a more accurate decision making, lower costs and reducing ATFM risk.

This paper primarily uses San Francisco International Airport (SFO) as a case study due to its uniqueness in employing two sources of terminal weather forecasts: Terminal Aerodrome Forecast (TAF) and the SFO Marine Stratus Forecast System (STRATUS). TAF is a weather forecast issued for all major airports and is updated four times in a day. It contains forecast information on visibility, ceiling etc for the entire day. Unlike other major airports, SFO experiences a low altitude marine stratus cloud layer during the summer which reduces the airport capacity. STRATUS forecasts the “burn-

off" time of these marine clouds i.e. the time when the capacity would increase. We construct probabilistic capacity profiles from both types of terminal forecasts.

The contribution of this paper is that it provides a methodology which uses several statistical techniques to convert weather forecasts into specific day-of-operation probabilistic capacity profiles. These profiles are provided as inputs in a Static Stochastic GDP model to simulate ATFM strategies. We have also developed a test-bed, assessing the benefit under the different methods of probabilistic profile generation against perfect information. The test-bed also provides an opportunity to score the benefits of the profiles generated by the two weather forecasts. The benefit on inclusion of the forecast is gauged by comparing the cost of the strategies developed with and without the weather forecast. Lastly, including the day-of-operation weather forecast would assist the air traffic controllers and dispatchers in planning a schedule for arrivals at the beginning of the day. The schedule could also be modified if new forecasts are available with time.

We have focused on the days when predictions from STRATUS were available i.e. marine clouds were observed in the terminal area. We construct probabilistic profiles from 7am to 10pm as the bulk of the traffic is observed in this duration. Since this paper deals with SFO, it may limit our generality of finding but this by no means is a limitation as the methodology can be applied to any other major airport for which the TAF is issued.

This paper proceeds as follows. Section II provides the relevant literature review. Section III describes the weather forecasts and the generation of probabilistic profiles by various methods. Section IV presents the GDP model and a comparative comparison of the costs of the strategies obtained from the profiles developed in Section III. Section V offers conclusion.

II. LITERATURE REVIEW

The current National Airspace System (NAS) rarely incorporates uncertainty of the weather forecasts into tactical decisions. Operations planning assume a deterministic approach using expected weather conditions [2]. Since it is difficult to accurately predict AAR, several researchers have formulated GDP models which require either stochastic capacity profiles or scenario trees for AAR as inputs [1,3,4]. This paper focuses on generation of stochastic profiles and details on scenario trees can be found in [5]. The generation of stochastic scenario found in the literature is primarily applied to stochastic programs in finance [6]. In the aviation community, a stochastic scenario for the capacity an airport refers to a set of several stochastic capacity profiles of the AAR. A stochastic profile is a time series of AAR values with a particular probability of occurrence. For a given airport there are several profiles depicting different possible evolution of capacity. Thus stochastic profiles capture the uncertainty in the future capacities.

Reference [5] formulates a methodology for developing stochastic profiles from historical AAR data for various airports in the United States. The profiles are the centroids of the clusters obtained after K-means clustering the AAR time

series. Their approach in profile construction is devoid of any weather forecast information. Reference [7] has focused on the STRATUS forecast thus its application seems to be limited to San Francisco International Airport (SFO). They assume a sharp capacity increase when the stratus "burns-off" which is not representative of the actual increase. They use the day of operation STRATUS forecast to determine the optimal end time of the GDP by modeling the time of the capacity increase as a random variable. Reference [8] gauged the imprecision of the forecast weather information with the actual weather by comparing delays. They compare the delays under the forecast conditions and actual weather conditions by associating a fractional loss of capacity under different weather conditions. Their approach ignores the imprecision in the forecast, which when implemented in ATFM simulation leads to a higher cost.

There is a gap in the research to develop specific day-of-operation capacity profiles or profiles based on weather forecasts.

III. WEATHER DATA AND PROFILE GENERATION

This section provides methodology for developing the set of probabilistic capacity profiles $\{S_p\}$ and their probabilities $\{P_p\}$ using various statistical techniques for both types of weather forecasts. With each technique a discussion is also provided to determine the number of profiles. An explanation of the relevant weather forecast precedes the development of profiles. Since we have focused on San Francisco International Airport, we also present a brief case study for the airport.

A. San Francisco International Airport

San Francisco International Airport is the second busiest airport in the California and the tenth busiest airport in the United States. It is also a major hub for United Airlines and a transfer terminal for international travel. The current runway configuration at SFO consists of four runways. The runways are positioned in parallel in two pairs. Two of them are in the north-south direction and the other two are the east-west direction. The separation between the parallel runways is 750ft. In good weather, simultaneous landings can take place on parallel runways accommodating a maximum arrival rate of 60 planes an hour. In bad weather, the landings are restricted to one runway which can reduce the capacity to 30 planes per hour. The 01R-19L is the smallest runway (2280m) while the 28R-10L is the airports' longest runway (3610m). When the 28 runways are in configuration, 28R is normally used for landings.

The San Francisco Bay and the Pacific Ocean are in close proximity to SFO and during the summer months marine stratus clouds (fog) are prevalent over the terminal area. During these months, the fog sets in by early morning and burns off by late morning or early afternoon. The marine stratus clouds play a critical role in determining the airport capacity. When the stratus clouds are below 3500ft in the vicinity of SFO, the airport operates in Instrumental Meteorological Conditions (IMC). The fog prohibits the use of the dual landing and the capacity of SFO is reduced leading to a lower AAR. If the fog coincides with the morning arrival push, it creates a demand capacity imbalance, formulation of en-route airspace queues and delays. On such days the Ground Delay Program is

initiated at SFO and its initial planned duration is the burn off time (or improved weather conditions) plus two hours. Due to the strategic importance of SFO and the effect of weather on operations there are two different weather forecast systems available. These predict different yet relevant, to aviation, meteorological conditions. Due to this unique nature of SFO we have used it as a case study. Both the forecasts are used to construct day-of-operation probabilistic profiles.

In the next subsections we describe the various methodologies for profile generation.

B. Naïve Clustering

This method of profile generation does not incorporate any weather forecast and is similar to [5]. In [5], the authors implement K-means clustering to generate the profiles by averaging the AAR time series in the same cluster. Let $[A]_{T \times N} = [A_1, A_2, \dots, A_N]$ be the data matrix of the AAR profiles for N days where A_k is column vector of the AAR profile for day k . K-means clustering splits the data matrix into a predefined number of clusters, I , where each cluster c_k contains d_k days. The days which have similar AAR profile vectors are grouped together i.e. they are in the same cluster. The similarity is defined as the Euclidean norm between the AAR profile vectors. A smaller value of the Euclidean norm implies greater similarity. After the K-means operation we obtain

$$\{A_h^1\}_{h=1}^{d_1}, \{A_h^2\}_{h=1}^{d_2}, \{A_h^3\}_{h=1}^{d_3}, \dots, \{A_h^l\}_{h=1}^{d_l}$$

such that $\sum_{j=1}^l d_j = N$ and

$$\bigcup_{j=1}^l \{A_h^j\}_{h=1}^{d_j} = \{A_h\}_{h=1}^N \text{ and } \bigcap_{j=1}^l \{A_h^j\}_{h=1}^{d_j} = \Phi \quad (1)$$

$\{\cdot\}$ is defined as a set

The optimal number of clusters, I^* , is an open problem and there are ad-hoc procedures which assist in determining it. More clusters enable tailor made ATFM strategies for each cluster but each cluster would have a lower probability of occurrence. Reference [5] provides an algorithm involving the pseudo F value to determine the optimal number of clusters.

Procedures like the pseudo F value and Silhouette value measure the compactness of a cluster with respect to other clusters and report an average value over all clusters. The pseudo F value is implemented in SAS and works well with uncorrelated variables (eg: variables obtained after Principle Component Analysis) [9]. The pseudo-F statistic captures the “tightness” of clusters, and is a ratio of the mean sum of squares between clusters to the mean sum of squares within a cluster. Higher pseudo F-values indicate tight clustering and imply that the data is well separated or better clustered. Silhouette value, varying between -1 and 1, measures the similarity between an object and the cluster in which it is classified. The indicator of a strong clustering is the average silhouette value close to 1 [10]. The procedure for silhouette value is implemented in MATLAB. Caution should be exercised in monitoring the number of data points falling

within each cluster. If they are too many data points within one cluster one might consider breaking it up on the other hand if there are too few days one would tend to merge two clusters together.

The profiles are the within cluster means of the AAR time series in that cluster. Profile S_i is determined by the average of the AAR profiles in the cluster c_i

$$S_i = \left[\frac{\sum_{h=1}^{d_i} A_h^i}{d_i} \right] \quad i \in 1..I^* \quad (2)$$

Where $[]$ is the nearest integer roundup operator.

The probability of the profile S_i is the proportion of days in c_i . $P_i = d_i/N \quad i \in 1..I^*$

For our data, $N=446$ days and $I^* = 3$. The number of clusters was determined by the highest Pseudo F value.

We call this procedure **Naïve Clustering** as it clusters the AAR without any weather information. Fig.1 shows the profiles for SFO for the summer months.

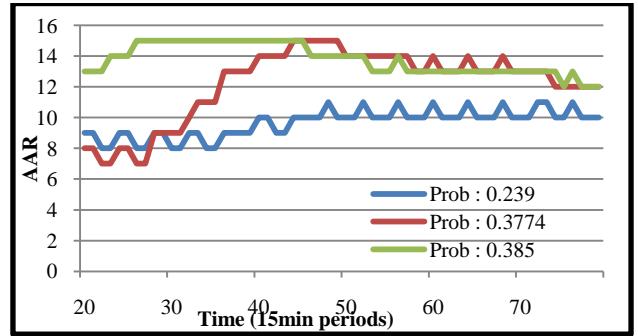


Figure 1. Stochastic Capacity Profiles from Naive Clustering

In the above figure, we observe an oscillating nature of individual profiles. This is because of the way the AAR is reported in the Aviation System Performance Metrics (ASPM) database. The AAR of 60/hour is split at 15,15,15,15/quarter hour, an AAR of 45/hour is split as 10,11,12,12/quarter hour and an AAR of 30/hour is split as 8,7,8,7/quarter hour leading to an oscillating nature of the profiles when reported in quarter hour intervals.

C. STRATUS and Fog Clustering

STRATUS is a program designed by MIT Lincoln Labs to forecast the time of the fog burn-off. It also includes the probability that the fog would burn off before 17Z, 18Z and 19Z (Z is UTC time) based on the empirical data of all burn-off time forecasts in the same 15 minute time bin. The probabilities are based on an ensemble of statistical models for the weather forecast and atmospheric boundary layer physics models [11]. STRATUS updates the forecast of the burn off time on an hourly basis from 2:00-11:00am PCT. NASA Ames Research Centre maintains a repository where the output from STRATUS is stored for the dates when marine clouds are forecast in the terminal area. For these dates, the data contains

the predicted burn off time, its actual burn off time and the probability that the fog would burn-off before 1700, 1800 and 1900 UTC Time.

We consider the STRATUS forecast generated at 8:00am PCT for over 200 days for the summer months (May - September) of 2004-2007. This was chosen as at 8:00am, predictions from the Satellite Statistical Forecast Model (SSFM) in the ensemble become available. We concentrated on the days when the fog burned off between 9:30am and 11:30am PCT as the number of days outside this time bracket were few. These days are binned in 15 minute periods according to the actual fog burn-off time between 9:30 to 11:30 a.m. PCT. In total there are eight fog burn off bins $\{B_k\}_{k=1}^8$. The number of days, d_i , in bin, B_i , is shown in the Table I. From each bin we constructed a probabilistic profile as follows:

$\{A_i^{B_k}\}_{i=1}^{d_k}$ is the set of AAR profiles for the days in bin B_k ($k \in 1, 2, \dots, 8$). The profile S_i is determined by the average of AAR profiles in B_i

$$S_i = \left[\sum_{h=1}^{d_i} A_h^{B_i} / d_i \right] i \in 1..8$$

(The above equation is similar to (2))

The profiles are shown in the Fig. 2.

TABLE I. BINNING OF ACTUAL FOG BURN-OFF TIME

Bin	Number of Days
9:30-9:45am	15
9:45-10:00am	16
10:00-10:15am	16
10:15-10:30am	11
10:30-10:45am	24
10:45-11:00am	22
11:00-11:15am	18
11:15-11:30am	15

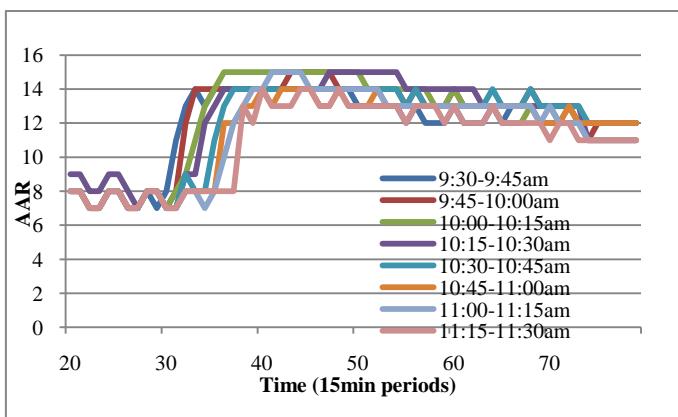


Figure 2. Profiles from Fog Clustering

A closer inspection of the periods when the fog burns off is shown in Fig. 3. It is seen that there is not an immediate increase in the AAR from 30/hour to 60/hour as assumed in [7]. There is a transition period lasting approximately for 45 minutes when the AAR is 45/hour. While calculating the ideal GDP end time, this transition should be taken into account. Ignoring this transition period would lead to an increased cost of airborne delays as the capacity would be over predicted immediately after burn-off. Fig. 3 shows the rise in AAR after burn-off.

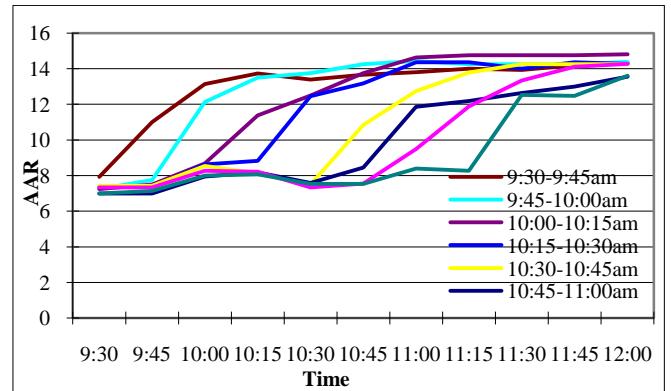


Figure 3. Rise in Capacity at Burn-off

MIT Lincoln labs, on recommendation by the Traffic Management Unit at Oakland center, incorporated “risks” to the output of STRATUS. The “risk” of an hour is the probability that the fog would burn off before that hour. The “risk” is thus a non decreasing function of time. The “risks” can determine the probability of the profiles. Since the risk probabilities are of the form $P(\text{Burn off} < T_1) = P_1$, $P(\text{Burn off} < T_2) = P_2$ and $P(\text{Burn off} < T_3) = P_3$ where $T_1 < T_2 < T_3$, we linearly interpolate the probabilities between the time periods to formulate a Cumulative Distribution Function (CDF) for the fog burn off time.

From the CDF the probability of any bin, B_i , can be calculated.

$$\overline{P}_{B_i} = \text{Prob}(\text{Burn off} \leq [B_i]) - \text{Prob}(\text{Burn off} \leq [B_i]) \quad i \in [1..8] \quad (3)$$

Where $[\cdot]$ and $[\cdot]$ are the lower and upper bin boundaries.

Equation (3) establishes the probability of the burn off in a particular bin. Further, if the burn off probability in a particular bin, B_i , is \overline{P}_{B_i} , then the capacity scenario, S_i , depicting burn off in B_i , would have a probability

$$P_i = \frac{\overline{P}_{B_i}}{\sum_{i=1}^8 \overline{P}_{B_i}} \quad \forall i \quad (4)$$

Equation (4) is a simple renormalization of the probabilities of the bins.

In conclusion, we have generated 8, 15 minute burn-off bins corresponding to the capacity profiles as shown in Fig. 2. From the STRATUS predictions of fog burn-off time for the

day-of-operation we can obtain the probabilities for the bins and consequently the probabilities of the profiles.

This methodology translates the STRATUS forecast to build probabilistic capacity profiles. We call this procedure **Fog clustering**.

D. TAF and Profiles from TAF

The Terminal Aerodrome Forecast (TAF) is a weather forecast issued for every major airport 4 times a day at 6 hours interval by the National Oceanic and Atmospheric Administration (NOAA). It contains the meteorological conditions (wind speed and direction, visibility and cloud type and height) along with qualitative descriptors (rain, fog, mist etc.) for the airport.

A script was written in MATLAB to parse through the data and select TAF data which was issued between 5am and 7am PCT. The TAF issued between this interval assists the SFO control tower, FAA and the airlines to plan the daily operations at SFO during the morning teleconference. In total, we had 446 days when the TAF was issued in this interval. As mentioned earlier the daily operations were considered between 0700-2200hrs PCT (i.e. 60 periods, 1 period = 15 mins). Another script in MATLAB parsed through this filtered data to pick out the relevant weather information: wind, visibility, clouds and their height (wind was broken in two components, wind from north and wind from east direction) for each period. Thus each day was represented by a column vector of length 60 (time periods) \times 7 variables (windNorth, windEast, visibility, 4 different clouds height) = 420. Therefore, the entire TAF data set could be represented by a 420 (variables) \times 446 (# of days) matrix.

Let $[T]_{L \times N}$ be a matrix ($N=446$, $L=420$), where T_k is a column representing the TAF for day k . We performed a Principal Component Analysis (PCA) on this matrix. PCA is a standard statistical technique which reduces the dimensionality of the data by converting correlated variables into a smaller number of uncorrelated variables called principal components. The principal components are directions which represent the variation in the data. Thus the first principal component direction represents the maximum variability in the data and each succeeding component accounts for as much of the remaining variability as possible. PCA removes the potential correlation between the forecasted variables for the same day [12]. For example, there might be correlation between visibility and ceiling and also there might be correlation between the forecast weather conditions of adjacent time periods. As a standard preprocessing technique, we normalize the $[T]$ matrix i.e. the mean and the variance is 0 and 1 respectively for each variable. Equations (5) through (9) describe the PCA on the data set.

$$[C]_{L \times L} = [T][T]^t / (N - 1) \quad (5)$$

Where $[C]$ is the empirical correlation matrix.

$$CX = \lambda X \quad (6)$$

Where λ is the eigenvalue corresponding to the eigenvector X

Sort the eigenvalues in a descending manner (matrix is full rank) i.e.

$$\lambda_{[1]} > \lambda_{[2]} > \lambda_{[3]} > \dots > \lambda_{[L]} \quad (7)$$

A standard technique is to capture 90% variability, the number of eigenvalues required are

$$n = \operatorname{argmax}_k \frac{\sum_{p=1}^k \lambda_{[p]}}{\sum_{p=1}^L \lambda_{[p]}} \leq 0.9 \quad (8)$$

Let eigenvector $X_{[i]}$ correspond to its eigenvalue $\lambda_{[i]}$, then

$$\text{define the matrix } [W]_{n \times L} = \begin{bmatrix} -X_{[1]}^T & - \\ \vdots & \\ -X_{[n]}^T & - \end{bmatrix}$$

The reduced TAF matrix

$$[\tilde{T}]_{n \times N} = [W] \times [T] \quad (9)$$

We proceed to perform a K-means clustering on the matrix $[\tilde{T}]$. It has been proved in [13] that performing PCA prior to K-means increases the accuracy of the K-means clustering. The centroids of the clusters are therefore closer to the optimal cluster centroids after PCA.

Thus a K-means clustering on $[\tilde{T}]$ with I predefined clusters leads to the following

$$\{\tilde{T}_h^1\}_{h=1}^{d_1}, \{\tilde{T}_h^2\}_{h=1}^{d_2}, \{\tilde{T}_h^3\}_{h=1}^{d_3}, \dots, \{\tilde{T}_h^l\}_{h=1}^{d_l}$$

such that $\sum_{j=1}^l d_j = N$ (10)

Where d_k is the number of days in the cluster c_k

$$\cup_{j=1}^l \{\tilde{T}_h^j\}_{h=1}^{d_j} = \{\tilde{T}_h\}_{h=1}^N \text{ and } \cap_{j=1}^l \{\tilde{T}_h^j\}_{h=1}^{d_j} = \Phi \quad (11)$$

From this analysis on the TAF issued for SFO, the optimal number of clusters were 2, $I^* = 2$ i.e. c_1 and c_2 . The number of clusters was determined using the maximum Psuedo F value. Thus, we classify a day in either c_1 or c_2 depending on the classification of its TAF.

We performed another K-means clustering on the AAR profiles of the days within the c_1 and c_2 . Let $\{\tilde{A}_h^1\}_{h=1}^{d_1}$ be the set of AAR profiles for the days in c_1 and $\{\tilde{A}_h^2\}_{h=1}^{d_2}$ be the set of AAR profiles for the days in c_2 .

The optimal number of AAR clusters within c_1 are 2 i.e. $c_{1,1}$ and $c_{1,2}$. Likewise, the optimal number of AAR clusters within c_2 are 3 i.e. $c_{2,1}$, $c_{2,2}$ and $c_{2,3}$. Thus $\{\tilde{A}_h^{1,1}\}_{h=1}^{d_{1,1}}$ and $\{\tilde{A}_h^{1,2}\}_{h=1}^{d_{1,2}}$ were in cluster c_1 and $\{\tilde{A}_h^{2,1}\}_{h=1}^{d_{2,1}}, \{\tilde{A}_h^{2,2}\}_{h=1}^{d_{2,2}}, \{\tilde{A}_h^{2,3}\}_{h=1}^{d_{2,3}}$ in cluster c_2 .

The highest average Silhouette value determined the number of clusters in this second K-means clustering.

The profiles and their probability are determined as below

$$s_{1,1} = \left[\sum_{h=1}^{d_{1,1}} \tilde{A}_h^{1,1} / d_{1,1} \right] \quad p_{1,1} = d_{1,1} / (d_{1,1} + d_{1,2}) \quad (12a)$$

$$S_{1,2} = \left[\sum_{h=1}^{d_{1,2}} \tilde{A}_h^{1,2} / d_{1,2} \right] \quad P_{1,2} = d_{1,2} / (d_{1,1} + d_{1,2}) \quad (12b)$$

$$S_{2,1} = \left[\sum_{h=1}^{d_{2,1}} \tilde{A}_h^{2,1} / d_{2,1} \right] \quad P_{2,1} = d_{2,1} / (d_{2,1} + d_{2,2} + d_{2,3}) \quad (12c)$$

$$S_{2,2} = \left[\sum_{h=1}^{d_{2,2}} \tilde{A}_h^{2,2} / d_{2,2} \right] \quad P_{2,2} = d_{2,2} / (d_{2,1} + d_{2,2} + d_{2,3}) \quad (12d)$$

$$S_{2,3} = \left[\sum_{h=1}^{d_{2,3}} \tilde{A}_h^{2,3} / d_{2,3} \right] \quad P_{2,3} = d_{2,3} / (d_{2,1} + d_{2,2} + d_{2,3}) \quad (12e)$$

Fig. 4 and Fig. 5 show the profiles and their probabilities for c_1 and c_2 .

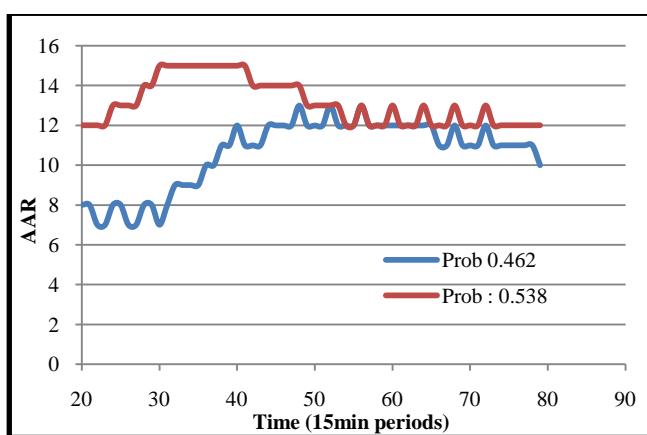


Figure 4. Probabilistic Profiles for cluster 1

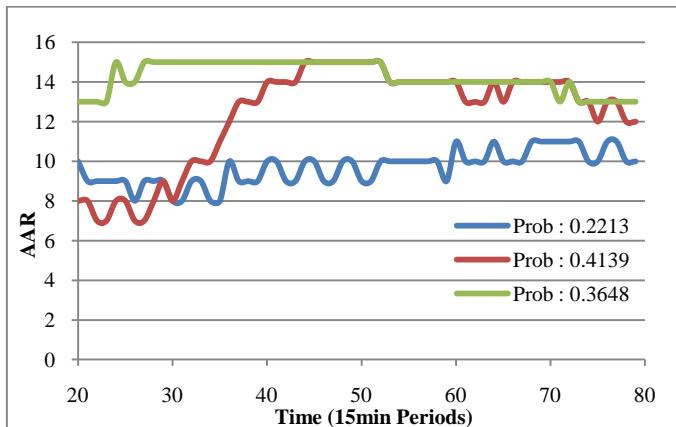


Figure 5. Probabilistic Profiles for cluster 2

We call this procedure **TAF Clustering**

E. Dynamic Time Warping Profiles

Dynamic Time Warping (DTW) is an established methodology to study the similarity between two electrical signals. It has been used in the areas of speech recognition to

match speech patterns. Reference [14] has demonstrated that DTW can be useful to detect similar multidimensional time series. DTW is a technique where one sequence is “warped” in time around the other. A distance matrix is generated between all the time pairs of the two series. A least expensive path in the distance matrix is determined dynamic programming. The path is subjected to certain end point constraints. We use the technique of dynamic time warping to compare the day of operations TAF with historic TAFs. AAR capacities of the historically similar TAF days are the actual capacity profiles and the probabilities of the profiles are proportional to the least expensive path length between the historically similar TAF with the day of operations TAF.

The multidimensional time series is the TAF for the day-of-operation and we have used the Euclidean norm to generate the distance matrix. Further we have restricted the search of the least expensive path to 3 periods as we want to search for similarity within a neighborhood of a period. Once we have the least expensive path lengths between the day of operations TAF and all the historic TAF days we can find the days which are most similar to the day of operation TAF. The mathematical formulation takes the following form.

Let \mathbf{T}_d be a time series of the TAF for the day of operation. \mathbf{T}_d is thus a 7 dimensional time series of length 60. The 7 dimensions represent the forecast in a period and the day-of-operation is divided in 60 periods.

Let $\{\mathbf{T}_i\}_{i=1}^N$ be a set of TAFs. DTW evaluates the least expensive path between \mathbf{T}_i and \mathbf{T}_j . A distance matrix of size 60×60 is first computed. Any element (r,s) of the distance matrix is $\mathbf{D}(r,s) = \|\mathbf{T}_i(r) - \mathbf{T}_j(s)\|_2^2$ ($r,s \in [1..60]$). $\mathbf{T}_i(r)$ is the TAF for period r for day i and similarly $\mathbf{T}_j(s)$ is the TAF for period s for day j . Thus a distance matrix for all possible day pairs is computed i.e. a total of $\frac{N(N-1)}{2}$ matrices are computed.

The DTW path between \mathbf{T}_i and \mathbf{T}_j is thus

$$\begin{aligned} DTW(\mathbf{T}_i, \mathbf{T}_j) = & D[\mathbf{T}_i(t), \mathbf{T}_j(t)] + \\ & \min\{DTW[\mathbf{T}_i(t-1), \mathbf{T}_j(T-1)], DTW[\mathbf{T}_i(t-1), \mathbf{T}_j(t)], DTW[\mathbf{T}_i(t), \mathbf{T}_j(t-1)]\} \quad (13) \end{aligned}$$

This path has to start from $t=1$ and has to end at $t=60$. These are the end points conditions.

So if two sequences, A and B, are completely similar i.e. identical, the $DTW(A,B)$ is 0. The optimal path would be the diagonal of the distance matrix.

In actual operations the day-of-operation TAF should be compared to the TAF of the days which precede it but due to data limitations we considered all the days. We experimented with the number of profiles following rule to determine the number of profiles:

Number of profiles for the day i with TAF \mathbf{T}_i

$$n = \operatorname{argmax}_k [\sum_{j=1}^k DTW(\mathbf{T}_i, \{\mathbf{T}_{[j]}\})] \leq \text{Const} \quad (14)$$

Const is a tuning parameter determined by experimentation with the number of scenarios. We found our solution to be relatively stable for the profiles between 8 to 12 and based on

this adopted $\text{Const} = 2500$. A lower value of Const would reduce n . If the TAF is accurate uncertainty in the AAR can be represented by a few profiles. If TAF is inaccurate then less scenarios would imply that we may be ignoring certain representations of the AAR which would increase the cost of delays.

Where $\{T[j]\}$ is an ordered set such that

$$\begin{aligned} DTW(T_i, T[1]) &\leq DTW(T_i, T[2]) \leq DTW(T_i, T[3]) \leq \dots \leq \\ &DTW(T_i, T[N]) \end{aligned} \quad (15)$$

The set of profiles is thus the actual AARs for the 'n' days.

$$S[k] = AAR[k] \quad (\forall k \in [1], \dots, [n])$$

The profile probabilities are obtained after normalizing the least resistant path for the 'n' days.

$$P[k] = \frac{1 - \left(\frac{DTW(T_i, T[k])}{\sum_{j=1}^n DTW(T_i, T[j])} \right)}{n-1} \quad (16)$$

We call this procedure as **DTW Profiles**.

IV. COMPARATIVE COMPARISONS

This section compares the cost of the delays after implementing the probabilistic capacity profiles developed in Section III in an ATFM simulation. The ATFM simulation is conducted using the GDP model developed in [3]. This model minimizes the total of cost of delay in a GDP by determining the arrival rate at the airport and the number of the aircraft subjected to ground delays. The model outputs the number of planes which can land in the absence of air holdings, the Planned Airport Arrival Rate (PAAR). As mentioned in the introduction, uncertainty of the AAR is captured by the probabilistic capacity profiles. In the model the cost of air delay c_a is assumed to be greater than c_g (otherwise there would not be a need to ground hold the aircraft). The model takes the following form:

$$\text{Min} \left[\sum_{t=1}^T c_g \times G(t) + \sum_{p=1}^N \sum_{t=1}^T c_a \times W(S_p, t) \times P_p \right] \quad (17)$$

Subject to:

$$\begin{aligned} A(t) - G(t-1) + G(t) &= D(t) \\ \{\forall t \in 1..T+1, G(0) = G(T+1) = 0\} \end{aligned} \quad (18)$$

$$\begin{aligned} -W(S_p, t-1) \pm W(S_p, t) - A(t) &\geq -M(S_p, t) \\ \{\forall t \in 1..T+1, -W(S_p, 0) = -W(S_p, T+1) = 0, \} \\ p \in 1..N \end{aligned} \quad (19)$$

$$A(t), W(S_p, t), G(t) \in \mathbb{Z}_+ \quad \{\forall t \in 1..T+1, p \in 1..N\} \quad (20)$$

Where, t : is the time period, S_p : p^{th} capacity profile (length T); P_p : is the probability of profile p ; T : total number of time periods or planning horizon; N : total number of profiles; $G(t)$: ground holding at time t ; $W(S_p, t)$: air holding under profile S_p at time t ; $A(t)$: planned airport acceptance rate at time t (**PAARs**); $M(S_p, t)$: capacity under profile S_p at time t ; $D(t)$: demand in period t ; c_a : cost of airborne delay; c_g : cost of ground delay; $\{S_p\}_{p=1}^N$ is the set of profiles; $\sum_{p=1}^N P_p = 1$

The objective function, (17), minimizes the sum of the fixed ground delay costs and the (expected) air delay costs. Equation (18), is a queuing constraint for flights bound for the destination from all the origin airports. It enforces flow conservation. The demand at period t , $D(t)$, plus the backlogged planes ground held in period $t-1$, $G(t-1)$, should either land, $A(t)$, or be put in a queue, $G(t)$. Equation (19) is a queuing constraint at the destination airport. Under capacity profile p , all planes that can land, $A(t)$, or air delayed from the previous time period $W(S_p, t-1)$ either land or are further air delayed to the next period, $W(S_p, t)$. The inequality is required as the demand might be less than the available capacity (runway + airspace). Equation (20) ensures that $A(t)$, $W(S_p, t)$ and $G(t)$ are real positive integers.

The decision variables are the number of aircrafts landing in a period t , $A(t)$, the number of aircrafts which are subjected to ground holding $G(t)$ and the number of aircrafts subjected to air borne holding under profile p , $W(S_p, t)$. The ratio of the cost of delays is selected to be 3:1:: $c_a:c_g$ based on heuristics and established conventions (the cost of air delays is roughly 3 times the cost of ground delay c_g is expensive due to higher fuel consumption in air, crew costs etc). The data for demand, $D(t)$ (planes originally scheduled to land in a period t) is obtained from the ASPM website. The model was solved in AMPL using the CPLEX as the solver with a run time of less than a second.

We simulated ATFM strategies for 50 historical days from 2004 to 2006 when the low lying marine stratus was observed at SFO and the results are based on these days. We generated the stochastic capacity profiles for each historical day using their TAF and STRATUS forecasts. For the Naïve case the profiles and probabilities were the same across all the days as the profiles are generated without weather forecast information. In case of the TAF Clustering, we first determined in which of the two clusters, c_1 or c_2 , the historical days TAF belong to and then applied the profiles and probabilities under that particular TAF cluster in the GDP model. So the uncertainty in the historical days AAR is represented by either 2 or 3 probabilistic capacity profiles depending on the classification of its TAF. Using the GDP model, we were able to compare the benefits under the different forecasts and different methodologies.

The GDP model determines four PAARs for the any historical day: PAARs_{fog}, PAARs_{TAF}, PAARs_{Naive} and PAARs_{DTW}. The PAARs are obtained using the probabilistic capacity profiles generated from Fog Clustering, TAF Clustering, Naïve Clustering and DTW Profiles methodologies respectively. To understand if the forecasts are useful in controlling the costs of delays, the four sets of PAARs are

compared in a deterministic queuing model. The idea for the deterministic queuing model is simple: the PAAR for a period, maybe higher than the actual realized capacity resulting in airborne delays from a capacity-PAAR imbalance. This also measures the similarity of the profile to the actual realized capacity. There are two types of delays if the PAAR is implemented 1) Airborne delays between the PAAR and the actual realized capacity 2) Ground delays between the PAAR and original schedule. The total cost (TC) of delay is $c_g \times \text{ground delay} + c_a \times \text{airborne delay}$. We obtain TC_{fog} , TC_{naive} , TC_{TAF} and TC_{DTW} from the queuing model.

The various total costs of delay are compared to a **Perfect Information** (PI) case where the controllers have perfect foresight about the evolution of capacity as if told by an “oracle”. For any historical day, we know the actual realized capacity and this capacity can be used in a (deterministic) ATFM simulation. This is equivalent to having one profile which is the actual AAR profile with 100% probability of occurrence in the GDP model. Thus if we have perfect information we can subject ground holding to all the aircrafts which would experience air borne delays under the original schedule. We translate all the possible air borne delays to ground delays, lowering total costs. Therefore, $TC_{\text{PI}} = \text{ground delay}$. TC_{PI} is the minimum possible cost that can be achieved for the day-of-operation. The average costs of delay for 50 days, obtained after the deterministic queuing model are shown in Table II.

TABLE II. AVERAGE COST OF DELAYS

Method	PI	DTW	TAF Clustering	Naïve	Stratus
Average	96.5	194.17	214.46	239.17	182
Std. Dev	54.9	125.89	145.28	156.72	109.93

From the average cost of delays it can be seen that average $TC_{\text{PI}} < \text{average } TC_{\text{fog}} < \text{average } TC_{\text{DTW}} < \text{average } TC_{\text{TAF}} < \text{average } TC_{\text{naive}}$. This illuminates the fact that probabilistic profiles derived from weather forecasts are better in planning of operations as compared to profiles developed devoid of any forecast information. The STRATUS forecast gives the minimum average cost of delays. This cost is marginally higher than the average delay from the DTW Profiles. The average costs of both STRATUS and DTW are almost double the average costs of PI. Days which have lower DTW costs than Fog clustering indicate that the TAF is accurate in predicting actual weather whereas STRATUS is not accurate on those days. Days when Naïve clustering gives a lower cost than DTW and Fog clustering indicate that both the forecasts were inaccurate. The costs of delays from the probabilistic capacity scenarios rely on the quality of the forecast, if the forecast is accurate in predicting the capacity, the delays would be minimized. The results indicate that TAF can be used to integrate weather with ATFM decision making for all airports.

V. CONCLUSIONS

In this paper we have demonstrated how to integrate the weather forecasts to generate probabilistic scenarios for the day

of the operation. We concentrated on the summer months (May-September) from 2004 to 2006. This is the first step towards the incorporation of weather forecast information in NEXTGEN. Though, we have considered SFO in our experiments, the methodologies can be applied to any airport. It is shown that incorporating the weather information to plan the day of operation arrivals leads to reduced costs. It is important to note that STRATUS is designed specifically for SFO and particularly for the days when there is a low lying stratus over the airport thus its application for all airports is limited. From our cost of delay calculations it can be seen that the difference of the costs between DTW Profiles and Fog Clustering is small, so the TAF offers similar benefit in decision making when incorporating forecast information to construct the probabilistic capacity profiles. Thus the DTW methodology can be applied to all airports.

REFERENCES

- [1] A.Mukherjee and M. Hansen, “A Dynamic Stochastic Model for the Single Airport Ground Holding Problem,” Transportation Science Vol. 41, No. 4, November 2007, pp. 444-456
- [2] David A. Clark, “Investigating a New Ground Delay Program Strategy for coping with SFO Stratus,” 89th AMS Annual Meeting ARAM Special Symposium on Weather - Air Traffic Phoenix, AZ, 11-15 January 2009
- [3] M. Ball O., R. Hoffman, A. R. Odoni and R. Rifkin, “ A Stochastic Integer Program with dual network structure and its application to the Ground-Holding Problem,” Oper. Res. 51 167–171. 2003
- [4] O Richetta and A. R. Odoni, “ Solving optimally the Static Ground Holding Policy Problem in air traffic control,” Transportation Science 27, pp 228–238, 1993
- [5] P. Liu, M. Hansen and A. Mukherjee, “Scenario-based air traffic flow management: From theory to practice,” Transportation Research Part B: Methodological, Volume 42, Issues 7-8, August 2008, Pages 685-702
- [6] M. Kaut and S.W Wallace, “Evaluation of Scenario-generation Methods for Stochastic Programming,” Stochastic Programming E-Print Series, 2003-14, <<http://hera.rz.hu-berlin.de/seps/>>.
- [7] L. Cook and B. Wood, “A Model for Determining Ground Delay Program Parameters using a probabilistic forecast of Stratus clearing,” Eighth USA/Europe Air Traffic Management R&D Seminar (ATM2009), Napa, United States 2009
- [8] A. Klein, R. Jehlen and D. Liang, “Weather Index With Queuing Component For National Airspace System Performance Assessment,” 7th USA/Europe ATM R&D Seminar(ATM2007), Barcelona, Spain, 2007.
- [9] SAS Institute, 2004 SAS Institute, 2004. SAS/STAT 9.1 User’s Guide. Cary, NC.
- [10] L. Kaufman and P. Rousseeuw, Finding Groups in Data: An introduction to Cluster Analysis, Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics, New York: Wiley, 1990
- [11] D.A. Clark et al., SFO Marine Stratus Forecast System Documentation, Project Report ATC-319, MIT Lincoln Laboratory, Lexington, MA, 2006.
- [12] I. Jolliffe, Principal Component Analysis, Springer Verlag, New York, 1986.
- [13] C. Ding , X. He, “K-means clustering via principal component analysis.”, Proceedings of the twenty-first international conference on Machine learning, p.29, July 04-08, 2004, Banff, Alberta, Canada
- [14] M. Vlachos, M. Hadjieleftheriou, D. Gunopulos and E. Keogh, “Indexing Multi-Dimensional Time-Series with Support for Multiple Distance Measures”, In Proc. of 9th International Conf. on Knowledge Discovery & Data Mining (SIGKDD), Washington, DC, 2003

Track 7

Future Concepts and Innovative Ideas

Dynamic Allocation and Benefit Assessment of NextGen Flow Corridors

Ali N. Zadeh

ali.zadeh@metronaviation.com

Metron Aviation, Inc.

Dulles, VA 20166

Arash Yousefi

arash.yousefi@metronaviation.com

Metron Aviation, Inc.

Dulles, VA 20166

Ali Tafazzoli

ali.tafazzoli@metronaviation.com

Metron Aviation, Inc.

Dulles, VA 20166

Abstract—A flow-based modeling approach is proposed to identify candidate airspace for high-density flow corridors. The input to the model is a set of projected user-preferred, wind optimal, and unconstrained 4D trajectories (4DT). We compute Velocity Vector Fields (VVF) in the 4D space-time and cluster the velocity vectors both in time and space to define flow of aircraft when they fly their preferred trajectories under high capacity conditions. A sliding time window is implemented to dynamically create and optimize corridors' coordinates based on the changes in preferred trajectories. From this process we compute a NAS-wide corridor network that mimics the dynamics of user preferred trajectories. In operational setting, flights will have the option of joining a corridor that is closest to their optimal trajectory. Using NAS-wide simulation, we asses the benefit of corridor network by comparing efficiency gained by joining the corridor network against extra distance traveled to join the network. We show that much of the overall corridors benefit may be gained by creating very few corridors.

I. INTRODUCTION

Recent advances in Communication Navigation and Surveillance (CNS) technologies are changing operating conditions of the Air Traffic Controller (ATC) [1], [2]. Today, with the use of advanced data links, one controller could be able to track an aircraft and communicate with a pilot all the way from origin to destination. Advanced navigation equipment and data links such as Airborne Separation Assurance System (ASAS), Automated Dependent Surveillance Broadcasting (ADS-B) and Cockpit Display of Traffic Information (CDTI) enables pilots for self-separation [3]. These new features provide flexibility for airspace designers to implement new classes of airspace that are capable of accommodating multiple times more traffic than the current airspace structures [4].

One of the new classes of airspace introduced within NextGen is Corridors-in-the-Sky [5]. These are regions of airspace, generally a long and narrow pathway, in which aircraft move in common direction or trajectory and along parallel lanes. Strictly speaking, by this definition, today's jet routes qualify as corridors with their capacity determined based on minimum separation distances in different dimensions. Corridors can be reserved airspace within which ATC provides neither separation services nor authorization to enter. Those two functions would overlap only in the corridor transitional areas, near the entry and egress points of the corridor.

The term corridor covers a range of operational concepts proposed by researchers such as High Volume Tube Shaped

Sectors (HTS)[6], tubes[7], [8], [9], [10], highways-in-the-sky[11], and Dynamic Multi-track Airways (DMAs)[12]. Similar to NextGen, the concept of "Freeway Airspace" has been proposed as part of European "Single Sky" as isolated airspace at high altitude with special rules that accomodates parts of core intercontinental traffic and parts of long haul domestic traffic in Europe [13].

While corridors do not have to be restricted to high altitudes, they find their most natural home in high-altitude airspace. Three of the prominent characteristics of NextGen corridors that would distinguish them from today's airways are:

- Allowance for multiple (parallel) lanes of traffic;
- Capitalization on advanced CNS technology to enable changes in methods of separation, such as self-separation that potentially reduces separation standards within the corridor, or enables flying in formation;
- Dynamic activation rules to add or remove corridor structures, as needed, throughout a day.

A well-designed corridor may reduce the airspace complexity and increase airspace capacity by minimizing interference from crossing traffic. In the corridor, there might be several parallel lanes to increase its capacity, breakdown lanes to accommodate avionics failures and passing lanes to accommodate aircraft with different performance characteristics operating in the same corridor. This benefits the corridor users directly by enabling them to use tighter separation between aircraft safely, boosting the capacity of the corridor and providing more flexibility in choice of slots for prospective corridor users. In spite of potential benefits by off-loading part of demand from conventional sectors, corridor implementation may create additional workload for controllers. Coordination of entering and exiting traffic into and out of the corridors may create additional task-load for controllers in conventional sectors. Additionally, presence of a segregated self-separated airspace in the middle of a sector may degrade controller situational awareness. Human-In-The-Loop (HITL) studies are underway to study these adverse effects [14].

A corridor can be static or dynamic. The corridor may be dynamically shifted to avoid severe weather or to take advantage of favorable winds. Moreover, it may only be utilized during certain times of the day or in response to certain triggers. In a dynamic corridor, the 3D trajectory of

its centerline is a function of time and the corridor's length may stretch or shrink during the day. Hence, the start and end of the corridor may change during the corridor's lifetime. This information must be fed to the Air Navigation Service Provider (ANSP) and be available for the aircraft flying that corridor. For instance, whenever a large number of flights are scheduled to be traveling from New York to Chicago within an hour, a pre-defined corridor might be turned ON and used to speed their journey, then deactivated for the rest of the day.

Yousefi et al.[6] performed statistical analysis of city-pair traffic and showed that 33 percent of the total scheduled flights are operated between about 10 percent of the city-pairs. These city-pairs were identified as backbones of the airspace system and it was concluded that increasing the capacity of these routes could significantly improve the total system capacity. Additionally authors modeled the air traffic as flow of a fluid, where aircraft are the particles of the fluid. The velocity vectors for small volumes of airspace were then calculated as the resultant velocity vectors for individual aircraft. Accordingly, vector fields of the fluid velocity were created. It was proposed that the analysis of vector fields' topology can be used to determine the geometry and location of potential corridors.

Sridhar et al. [8] used clustering techniques using today's traffic to group the airports in close proximity of each other and construct corridors by connecting these groups of airports via great circle routes. Xue et al. [9] used Hough transformation to cluster the great circle routes between city-pairs into corridors and performed analysis of NAS-wide deviation from great circle routes required to join the corridor network. Hoffman et al.[10] constructed a corridor network by connecting major metroplex airports. Finally, Wing et al. [12] performed regional pooling of airports using different distance criteria and constructed DMAs by connecting pools of airports via great-circle routes.

In this paper, we attempt to analyze the dynamics of corridor network and include this in our design methodology. Specifically, we predict the periods that corridors should be active or how their centerline should dynamically change in response to changes in demand profile and weather disruptions. We develop a method that dynamically identifies high-density sections of the airspace that can benefit from new corridors. Similar to the proposal by Yousefi et al.,[6] we model the aircraft flow and then group the major traffic flows based on a predetermined set of proximity parameters and finally insert corridors along the flows' center of gravity. Furthermore, through NAS-wide simulation of traffic and using a futuristic 2.0X traffic forecast, we perform benefit assessment of corridor network in terms of system-wide recovered delay.

The paper is organized as follows: Section II, presents the steps and the mathematical model for designing a dynamic corridor backbone over a desired section of airspace. Section III discusses the cost/benefit analysis in terms of the total length of corridors implemented and the percentage of NAS-wide delay recovered. Finally, Section IV summarizes conclusion, briefly describes the possible implementation scenarios, and

indicates next research steps.

II. CORRIDOR ALLOCATION METHODOLOGY

The following are the design criteria we seek to address in our methodology:

- A corridor is expected to be the principle and best route between two en route points and therefore its location should roughly align with user-preferred and wind-optimized trajectories.
- We assume that aircraft inside corridors do not observe en route delay. This may be possible if we design corridors with sufficient lanes to accommodate the assigned traffic.
- Technically the more corridor we place in the system, the more delay will be recovered. However, in some areas the traffic density is not high enough to justify placement of corridors. Hence, projected corridor utilization must be considered as a design parameter.
- Only aircraft that are equipped with advanced avionics for self-separation are allowed to enter the corridor network.
- Corridor coordinates should be dynamic in time and an activation rule should be defined based on projected traffic density.

Using these criteria we first determine the optimal location of corridors and update their location based on the changes in user preferred trajectories. The following subsections describe the steps of our metrology.

As a final step, our algorithm suggests the best trajectory (including inside and outside a collection of corridors) that an aircraft could fly to observe minimum en route delay.

The following subsections describe the steps in our algorithm to design corridor backbone.

A. User-Preferred Trajectories and Traffic Forecast

Our main design criterion is to place the corridors close to user-preferred trajectories. Hence the first step is to generate futuristic unrestricted 4DTs including equipage forecast. The equipage data is needed to indicate which aircraft are capable for self-separation. We used a schedule data that was evolved from a baseline schedule of 3/28/2007 to 2.0X using FAA's 2007 Terminal Area Forecast (TAF) reports. This data was fed to NASA's Future Airspace Concept Evaluation Tool (FACET) to calculate wind-optimal trajectories. Finally, equipage data was fused to these trajectories and this unconstrained data is used as input to our analysis. Details of this data preparation process can be found in the cited references[15].

B. Flow Modeling based on Velocity Vector Field

In this section we develop a flow-based approach to identify candidate airspace for corridors. The space-time comprising of the earth sphere plus its exterior and time is a 4-manifold with boundary denoted by M . Let Tr denote all the aircraft trajectories in the 4-manifold M for an indefinite period of time. To each point on Tr , a tangent velocity vector is associated that follows the direction of the trajectory and indicates the velocity of the aircraft on that trajectory. For locations with no trajectory passing through them a zero vector

is assigned. Note that in reality, at each point in space-time there can be only one aircraft and therefore only one velocity vector.

We explain below that the proposed algorithm works even if some trajectories calculated from flight plans coincide at some points. However, to establish a sound mathematical background for this problem, it is assumed that any absolute conflict in trajectories is resolved beforehand. This can be achieved by trajectory optimization and speed regulation [16].

Equivalently, for a flight i at time t , the 4D position vector $r_i(t)$ and 4D velocity vector $v_i(t)$ are related as follows:

$$\vec{r}_i(t) = (x_i(t), y_i(t), z_i(t), t) \quad (1)$$

$$\vec{v}_i(t) = \left(\frac{\delta x_i(t)}{\delta t}, \frac{\delta y_i(t)}{\delta t}, \frac{\delta z_i(t)}{\delta t}, 1 \right) \quad (2)$$

where x_i , y_i , and z_i denote the aircraft's position components in Cartesian coordinate system.

The above characterization is the definition of a VVF on the 4-manifold M . Lambert conformal conic projection is used to create a vector space of state vectors.

C. Discretization of Velocity Vector Field

As a first step to make data suitable for numerical analysis, the user-preferred flight trajectories, Tr are sampled every T_s units of time starting from $t = 0$. The discretized version of the VVF is obtained by the set of velocity vectors defined for these sample points. Selection of T_s is based on the trade off between computational speed and accuracy of the results.

Each sampled data item includes four positions and three non-trivial velocities. These seven attributes are sufficient to create a VVF over 4D space-time. By presenting the magnitude of each velocity vector in terms of the distance the associated flight travels per sampling period T , we can immediately identify the flight trajectories in the VVF.

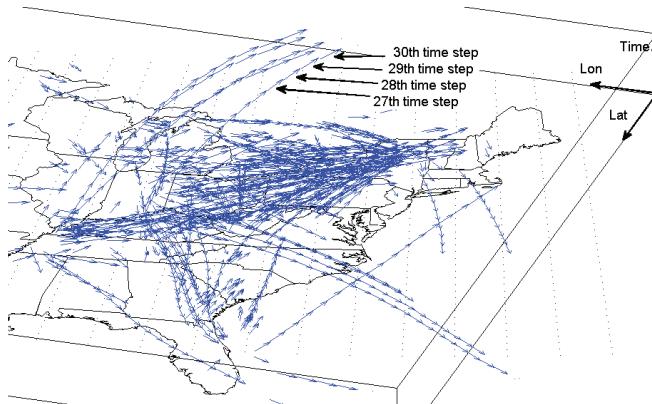


Fig. 1. VVF for FL280 to FL310 from 00:00 to 03:00 Zulu in 6-Minute Increments.

Figure 1 demonstrates a sample projection of VVF over the latitude-longitude-time space for three hours of sample data from 00:00:00 Zulu over the continental US. Only aircraft flying at flight levels FL280 to FL310 are included. As shown in the figure, for a high sampling rate, the resulting graph will be a close approximation of trajectories.

D. Clustering Velocity Vectors

Each non-zero velocity vector in the VVF is identified by the seven-element state vector $\mathbf{s} = (t, r, \theta, \phi, v, \alpha, \gamma)$ representing the sampling time, the position of its origin, and its speed and heading. t is the sampling time measured in number of elapsed time steps T from a reference point. r, θ and ϕ identify altitude, latitude, and longitude respectively. v identifies the ground speed, and α and γ denote the horizontal and vertical headings of the aircraft. This state vector is augmented by a weight factor ω that will carry some information from the clustering process. All the state vectors in the VVF are originally assigned a weight of 1. The heart of our methodology is to identify the state vectors that are adjacent and cluster them to one corridor element. The following definitions and procedures provide the bases for our algorithm.

Definition 1: A norm operation $\|\cdot\|$ on any two vectors \mathbf{s}_1 and \mathbf{s}_2 returns a non-negative real number that is a measure of their closeness. The following two norm operators $\|\cdot\|_1$ and $\|\cdot\|_2$ are used in different stages of corridor design procedure:

$$\|\mathbf{s}_1 - \mathbf{s}_2\|_1 = \max \left(\frac{|t_1 - t_2|}{T + 1/2}, \frac{|r_1 - r_2|}{R}, \frac{gcd(\theta_1, \phi_1, \theta_2, \phi_2)}{D}, \frac{|v_1 - v_2|}{V}, \frac{|\alpha_1 - \alpha_2|}{A}, \frac{|\gamma_1 - \gamma_2|}{\Gamma} \right) \quad (3)$$

$$\|\mathbf{s}_1 - \mathbf{s}_2\|_2 = \max \left(\frac{|t_1 - t_2|}{T + 1/2}, \frac{|r_1 - r_2|}{R}, \frac{latD}{\Theta}, \frac{lonD}{\Phi}, \frac{|v_1 - v_2|}{V}, \frac{|\alpha_1 - \alpha_2|}{A}, \frac{|\gamma_1 - \gamma_2|}{\Gamma} \right) \quad (4)$$

gcd is the great circle distance between two points on the surface of earth. $latD$ and $lonD$ calculate the lateral and longitudinal distances of two velocity vectors with respect to their mean heading. Variables $T, R, D, \Theta, \Phi, V, A$, and Γ specify separation requirement and are determined based on design criteria. $\|\cdot\|_1$ is used with $T = 0$, when we are grouping the vectors at each time step. $\|\cdot\|_2$ is used when we are grouping the elements of potential corridors in the sliding window protocol described in the next section.

Definition 2: Two state vectors \mathbf{s}_1 and \mathbf{s}_2 are considered neighbors if and only if:

$$\|\mathbf{s}_1 - \mathbf{s}_2\| \leq 1 \quad (5)$$

Definition 3: A set of state vectors is called a bundle and is identified by B .

Definition 4: The neighbor list of a state vector \mathbf{s}_i , denoted by NL_i , is defined as:

$$NL_i = \{\mathbf{s}_k : \|\mathbf{s}_i - \mathbf{s}_k\| \leq 1\} \quad (6)$$

Note that state vector \mathbf{s}_i is included in its own neighbor list.

Definition 5: A complete neighbor list CNL is a set of state vectors that is the union of all the neighbor lists of its

members:

$$CNL = \{s_k : \forall s_i \in NL_k \Rightarrow s_i \in CNL\} \quad (7)$$

Generating CNL is an iterative procedure of searching for the neighbors of the current members and adding the newly found state vectors to the bundle until all the members have been tried.

Procedure 1: Generating the CNL_k around a state vector s_k , known as the seed of the bundle:

- 1) Initialize CNL_k with the state vector s_k
- 2) For each $s_i \in CNL_k$ calculate the following:
 - Generate NL_i
 - Update CNL_k by concatenating it with the set $NL_i - CNL_k$

Proposition 1: The following condition always holds:

$$\text{if } s_i \in CNL_k \Rightarrow CNL_k = CNL_i \quad (8)$$

Procedure 2: Partitioning a set of state vectors S in a 4D section in space-time:

- 1) Maintain a check list Cl for all members of S
- 2) For each $s_k \in S$ calculate the following:
 - If $Cl(s_k) = 1$ continue with the next member
 - Otherwise generate CNL_k
 - Set $Cl(s_i) = 1$ for all $s_i \in CNL_k$
- 3) Proposition 1 ensures that the outcome of partitioning is indifferent to the initial seed location, and that the intersection of any two distinguished $CNLs$ is empty:

$$CNL_i \cap CNL_k = \emptyset \Leftrightarrow CNL_i \neq CNL_k \quad (9)$$

Definition 6: The Center of Gravity (CoG) of a bundle B is a state vector that is the weighted average of all its members:

$$s_{CG} = \frac{1}{\omega_{CG}} \sum_k \omega_k s_k, \quad (10)$$

where k is the index of all $s_k \in B$, ω_k is the weight of each state vector, and $\omega_{CG} = \sum_k \omega_k$ is the weight associated with CoG vector.

Definition 7: A CoG bundle is a set of state vectors that all its members are in the neighbor list of its CoG vector:

$$B_{CG} = \{s_k : \|s_k - s_{CG}\| \leq 1\} \quad (11)$$

Proposition 2: The CoG of two disjoint bundles B_1 and B_2 can be obtained by:

$$s_{CG} = \frac{1}{\omega_{CG1} + \omega_{CG2}} (\omega_{CG1} s_{CG1} + \omega_{CG2} s_{CG2}) \quad (12)$$

where ω_{CG1} and ω_{CG2} are the weights of the two bundles respectively.

Procedure 3: Generating a CoG bundle around state vector s_k :

- 1) Initialize CoG_k with the state vector s_k
- 2) Calculate the CoG vector s_{CG} for the set
- 3) Generate neighbor list NL_{CG} for s_{CG}
- 4) If $NL_{CG} - CoG_k = \emptyset$ then terminate the procedure

- 5) Otherwise update CoG_k by concatenating it with the set $NL_{CG} - CoG_k$
- 6) Go back to step 2.

Procedure 4: Breaking a complete neighbor list CNL to a set of CoG bundles:

- 1) Calculate CoG vector s_{CG} for the set
- 2) Order the members of CNL in descending order by their distance from s_{CG}
- 3) Pick s_k from the head of CNL
 - a) Calculate the CoG_k for state vector s_k
 - b) Remove CoG_k from CNL
 - c) Go back to step 1 until $CNL = \emptyset$

Moving from CNL to CoG bundle imposes a bound on maximum width of corridors.

Figure 2 demonstrates the steps in calculating CoG vector at each time step for two adjacent flights.

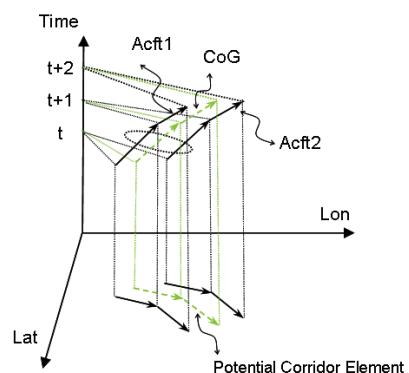


Fig. 2. Clustering Procedure.

E. Sliding Window Framework

We are now equipped with necessary structure and tools to design corridors element by element by clustering the aircraft state vectors. We first identify areas in the space-time with high level of congestion. Due to random nature of the NAS and uncertainties related to weather disruptions, the actual 4D trajectories are often diverted from the filed flight plans. If all the probability density functions (pdf) of the random variables are known, it is possible to obtain the chance of separation violation for any two aircraft. In this paper, we do not intend to obtain pdf for deviation of each aircraft from its centerline. Instead to account for the underlying uncertainty, we consider multiple times minimum required separation to identify the regions with potential violation and cluster the state vectors accordingly.

The second step in the design procedure is to calculate dynamically the attributes of each corridor element as listed below:

- Coordinates of the corridor's centerline
- Number of parallel lanes of traffic
- Activation time period
- Aircraft speed range

Assume at time step t a corridor network already exists. Consider a sliding time window of size T_{sw} starting at the processing time step t . The goal is to update the corridor backbone by adding additional corridor elements or removing or modifying existing corridors elements based on the preferred flight trajectories in this time window. If a corridor element is not highly occupied for some T_{ka} period of time, it becomes inactive. T_{ka} is known as the keep alive time.

When the new state vectors that are entering the sliding time window have slight deviation from some corridor elements, it is not possible to know if the deviation is in response to weather parameters, or simply due to association of the flight with a new origin destination pair. Therefore to create a stable corridors, clustering and CoG calculation is done according to the weighting of each element. Each corridor element is weighted based on the average flow rate it is accommodating.

We must also distinguish between desired 4DT and planned 4DT. At any time step, each aircraft files its desired optimal 4DT in the unconstrained airspace and the optimal 4DT that it plans to fly considering the structure of the corridor backbone and other constraints. In other words, depending on the location of the corridors more flight may change their planned 4DT to benefit from those corridors, however, the optimal design of corridors is done according to the VVF obtained from the desired 4DT of all the aircraft.

Corridor network is designed in the following high-level steps:

- 1) Create the VVF from the desired 4DT of all the aircraft
- 2) At a time step t , collect the state vectors for T_{sw} units of time in the future.
- 3) For each time step in the sliding window, calculate all the distinguishable *CNLs* with time separation parameter set to zero $T = 0$ and norm definition $\|\cdot\|_1$. The other separation parameters are design parameters that must be chosen for optimum cost/benefit tradeoffs.
- 4) Break the *CNLs* to the minimum set *CoG* bundles and calculate the flow rate for each bundle.
- 5) Replace all the *CoG* bundles with their associated *CoG* vector s_{CG} .
- 6) In the new VVF, for all the data in the sliding window, calculate all the distinguishable *CNLs* using the norm definition $\|\cdot\|_2$. Set the time separation parameter to the size of sliding window $T = T_{sw}$, this is equivalent to collapsing the 4D potential corridor elements over the time domain. Also, set the lateral separation Θ to distance separation D in $\|\cdot\|_1$. The longitudinal separation parameter Φ is set to half of the average speed of any two state vectors. The other parameters are the same as in $\|\cdot\|_1$.
- 7) Break the new *CNLs* to the minimum set *CoG* bundles and calculate the flow rate for each bundle. The set of associated *CoG* vectors s_{CG} with high usage over time are good candidates for permanent 3D corridors.
- 8) Filter out the low-flow rate elements based on the design parameter flow-rate threshold FR_T .
- 9) Update the corridor backbone with the remaining corri-

dor elements:

- a) For each *CoG* vector s_{CG} from the sliding window, search in the corridor backbone for a neighboring corridor element s_{CB}
- b) Replace s_{CB} with the *CoG* vector of the two adjacent vectors
- c) If there was no neighbor, add the new *CoG* bundle to the corridor backbone

10) Backbone structure is broadcast back to all aircraft.

11) Set $t = t + 1$ and go back to step 2.

By grouping the state vectors inside the sliding window, we can calculate the average traffic flow per unit of time. Moreover, the minimum required longitudinal separation of aircraft in a corridor determines the capacity of each lane; therefore the number of required lanes in each corridor element can be calculated based on the average flow rate in the corridor.

Note that depending on the accuracy of the weather forecast, the time-specific corridor elements may be published a few hours ahead of their activation.

F. Deployment Scenarios

To obtain the optimal corridors, we assume that each flight would file a desired unconstrained trajectory and an actual planned trajectory which is calculated based on the current structure of airspace and its associated constraints. Planned trajectory may divert from the optimal route because it may benefit from an existing corridor. However, corridor design in each time step, must be calculated according to their desired route.

Based on the 7 separation parameters, we can create an overlapping grid up to 7 dimensions. The goal is to minimize the search computation for finding neighbors to each state vector. The size of each dimension of the grid cell is multiple number of the separation parameter for that dimension. In the first step, each state vector is mapped to one of these grid cells, based on its parameters. To find the neighbors, we only search for other state vectors within that cell and its neighboring cells.

G. Navigational Reference System Grid

The FAA has proposed the Navigational Reference System (NRS) as part of the high-altitude airspace redesign initiative [17]. The NRS is a set of waypoints located on a regular grid of latitude/longitude coordinates, and is used for flight planning and navigation. We assume that within NextGen the NRS grid will be used as a referencing system and corridors should be published as a collection of consecutive NRS waypoints. A calculated corridor element based on *CoG* bundle does not necessarily fall on an NRS arc. Therefore, after generating the corridor network backbone, we map each corridor segment to its closest link in NRS grid.

III. RESULTS AND DISCUSSIONS

In this section we use a sample traffic forecast and generate corridor networks using the defined processes. Furthermore, we assess the trad-off between NAS-wide delay reduction as a result of employing efficient corridors, against the extra distance flown to join the corridor network.

A. Corridor Network Generation

We use the unconstrained, user preferred, and wind-optimal 4DTs generated in Section II-A and generate corridor backbone. Figure 3 depicts the time step clustering of VVF for one hour of flight schedule at 15 Hour Zulu time for all aircraft flying Continental United States (CONUS) at flight levels FL290 and above. For clustering the state vectors, separation parameters are set to: ($T = 0$, $R = \infty$, $D = 32\text{nm}$, $\Theta = 5^\circ$, $\Phi = 5^\circ$, $V = 30 \text{ Knots}$, $A = 5^\circ$, $\Gamma = \infty$). We have not set any separation limit in altitude. As depicted by the figure, velocity vectors almost cover the entire airspace but many of these routes are not flown frequently. Aircraft flying the same origin/destination but over different time periods may fly completely different routes that are optimized for wind-data at that specific time of the day. Note that some trajectories associated with different Origin/Destination may be merged for part of their path when they get close enough.

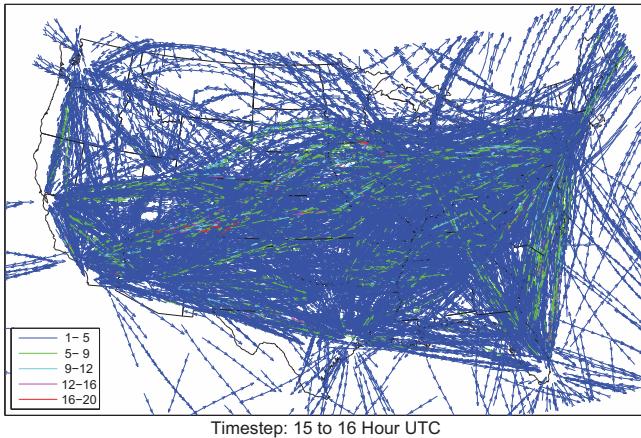


Fig. 3. Time Step Clustered Velocity Vectors for FL290 and Above.

In Fig. 4, the 4D VVF has been collapsed over its time dimension to obtain the potential corridor elements, and calculate the percentage of time each corridor element is occupied over time. In Fig. 5 the corridor elements have been mapped to NRS grid and data is filtered to exclude the corridor elements that have been occupied by flow rates of less than 10 aircraft per hour. Now the goal is to identify the potential 4D corridor elements that are highly utilized over time.

Time usage of potential corridor backbone has been illustrated in Fig. 6. Utility is calculated as the percentage of aircraft flying the corridors over time. FR_T is set at 10 aircraft/hour. As the flow rate is increased, the percentage of aircraft that are naturally in the corridors reduces. Figure 7 illustrates the scenario that the FR_T has been raised to 25 aircraft per hour and corridors are mapped to NRS grid.

It is apparent from these figures that higher values for FR_T result in less dense corridor network. Figure 8 illustrates the percentage of active corridors for different values of FR_T . For example for $FR_T = 25$, total length of corridor backbone is reduced to only 52,000nm or roughly 10 coast-to-coast corridors. This indicates the fact that there are certain flows of

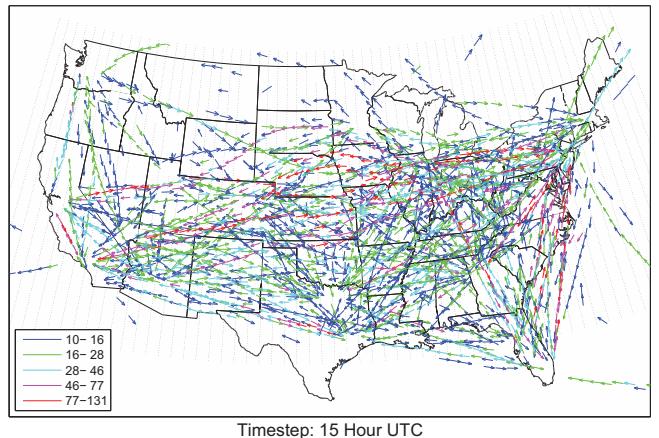


Fig. 4. Corridor Backbone for FL290 and Above with $FR_T = 10$.

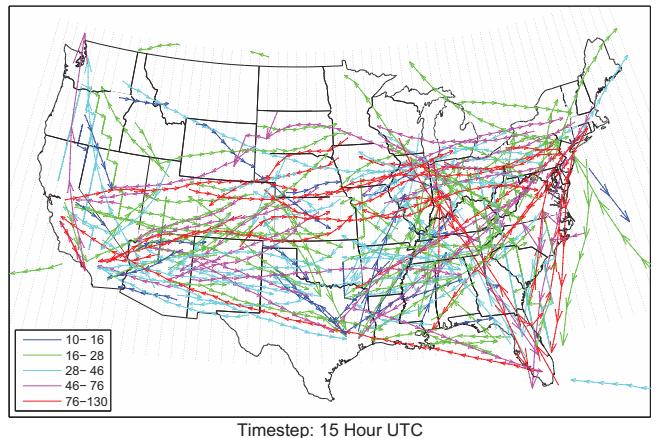


Fig. 5. Corridor Backbone for FL290 and Above with $FR_T = 10$.

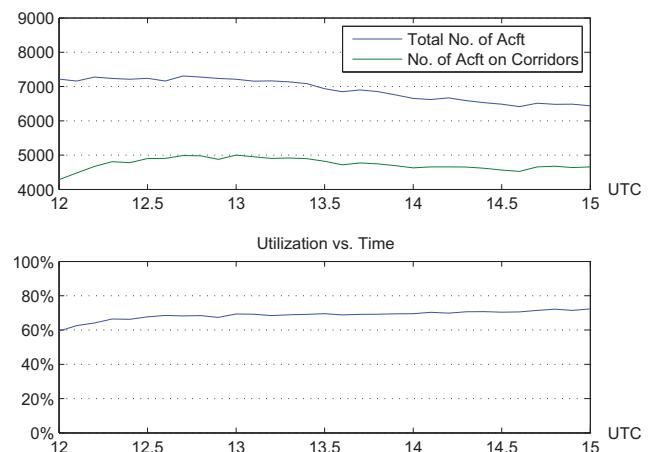


Fig. 6. Corridor Utilization Over Time for FL290 and Above with $FR_T = 10$.

aircraft in the NAS that carry a large portion of overall traffic and enhancing the en route efficiency of these flows (i.e. by employing corridors) would result in significant system-wide throughput enhancement. Next we show that creating corridors along these major flows may deliver much of the overall NAS-

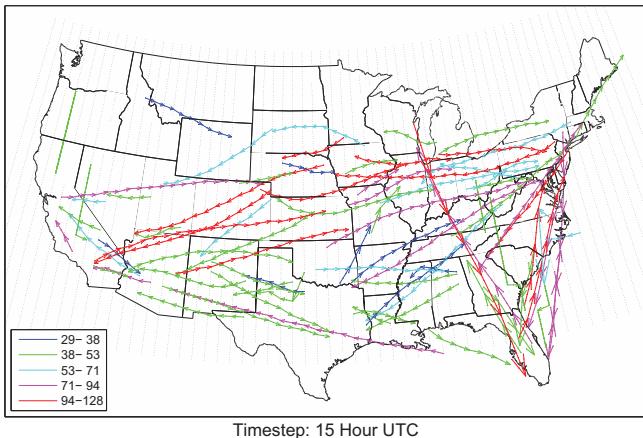


Fig. 7. Corridor Backbone for FL290 and Above with $FR_T = 25$. wide delay reduction.

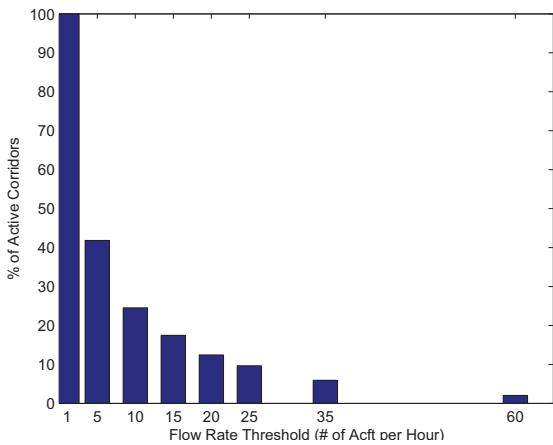


Fig. 8. Percentage of Active Corridors vs. Flow Rate Threshold.

B. Corridor Benefit Assessment

In this section we describe a method for evaluating the benefit of corridors in reducing NAS-wide delay. Consider two routing options between two en route points **A** and **B**, as shown in Fig.9. Aircraft can fly the direct route between two points and observe delay as a result of sectors' limited capacity; $delay(A, B) = \sum_{j=1}^5 d_j$, where d_j is the en route delay observed in sector j . Alternatively, aircraft can fly extra distance to join and leave a corridor element and therefore observe delay; $delay(A, B) = \sum_{j=1}^2 d'_j$, where d'_j is the delay due to excess distance flown. Comparison of these two delay values indicates the optimal trajectory for each flight and whether it is efficient for them to join the corridors. This assumes that corridors have enough capacity to allow aircraft to fly their optimal cruising speed.

NAS Simulation and Queuing Model (NSQM) is used to estimate the en route delay. NSQM is a NAS-wide discrete time simulation model developed to provide a level playing field for evaluating and comparing Traffic Management Initiatives (TMI) strategies [18]. It uses predetermined sector capacity

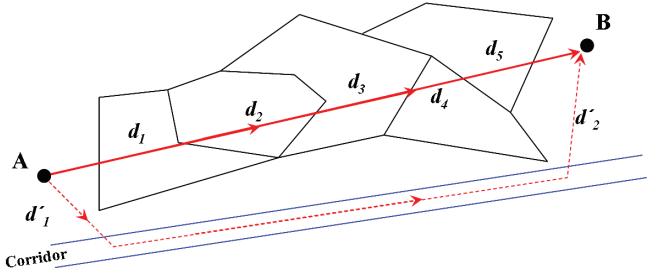


Fig. 9. Aircraft flying from **A** to **B** have the option of crossing multiple sectors or travel extra distances to join a corridor element.

values and applies airport holding, airborne delay, and re routes to maintain sectors' demand under admitted capacities. We define sector capacities based on historical usage; for additional information on how nominal sector capacities are defined refer to Myers et. al. (2008)[18].

The unconstrained wind-optimal shortest paths for a set of scheduled flights for the selected date are fed as the original flight plans to NSQM to obtain a set of sector inefficiency factors, f_{si} . f_{si} is calculated for each sector as the ratio of total delay incurred to total dwell time for all the flights going through that sector. This inefficiency factor indicates the average delay that each flight observes per unit of dwell time while crossing each sector. Finally, having these inefficiency factors for each sector, we can compute the delay that each sector will impose on each flight without a need for further NSQM runs.

For different values of flow rate threshold, FR_T , we generate the corridors and map them on NRS grid. A network graph is mapped over the NRS grid and the cost of each link is set as the physical length of the link divided by the aircraft ground speed. If the link is not overlapped with a corridor element, the link cost is then increased according to the inefficiency factor of the embedding sector. The optimal trajectory for each aircraft is now determined by obtaining the shortest path on the associated graph. In the final step, we compute NAS-wide delay as a function of total corridor length. The results of this analysis is summarized in Fig. 10. For different corridor lengths, the recovered NAS-wide delay is plotted as percentage of delay reduction from baseline no-corridor case. Higher values for FR_T result in less dense backbone of corridors. For example for $FR_T = 25$, total length of corridor backbone is reduced to only 52,000nm or roughly 10 coast-to-coast corridors. For this value, still about 60% of aircraft are inside the corridors at any time step and delay is reduced by about 60%. From this point, as we create more corridors (by decreasing FR_T), the rate of increase in recovered delay declines. This is an important observation and indicates that *much of the overall corridor benefit can be gained by creating few corridors*.

Publishing the corridors and deploying new procedures for interaction of corridor and non-corridor traffic is a costly exercise for the ANSP, especially when corridors are changing dynamically. However our analysis shows that few corridors should deliver significant benefit and this may justify creation

of few major corridors within NextGen.

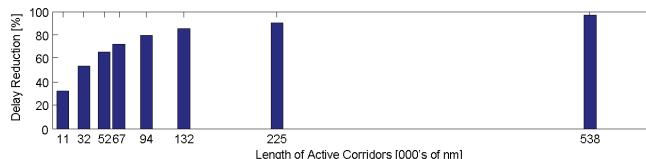


Fig. 10. Density of Potential Corridor Elements for Flight Level FL380.

IV. CONCLUSION AND FUTURE WORK

The issue of ATC workload as a critical capacity constraint is apparent. Without a revolutionary change, the ATM system will not efficiently handle the future growth in air traffic. Recent advances in avionics and data links provide capabilities for new concepts of operation. Corridors-in-the-Sky is one such concept that is proposed within the NextGen ConOps.

Objective methodologies are needed to dynamically compute the topology of the corridors. We have modeled the air traffic similar to flow of a fluid, where aircraft are the particles of the fluid. Contiguous corridors are created by connecting the high density and usage corridor elements which indicate high flow rate traffic. Then we map the resulting corridors to NRS grids. Using a set of unconstrained 4DTs, we presented sample corridor networks for different values of corridor flow rate threshold. We used detailed NAS-wide simulation to assess the benefit of corridor network in recovering overall airborne delay. We show that much of the overall corridors benefit may be observed by creating very few corridors in the NAS. Due to many parameters affecting the user-preferred routes, we believe that the optimal design of corridors and scheduling of flight plans should be performed iteratively. Unconstrained user preferred 4DTs may be fed to our algorithms to produce a corridor network. Then aircraft may decide to alter their routes to join some segments of corridor network. This new flight plan is passed to corridor design algorithm to update the corridors. This procedure should be iterated until an optimal corridor backbone and optimal flight plans are obtained.

In our methodology we define several separation parameters for clustering 4DTs into potential corridors. Additional work is needed to analyze the sensitivity of final corridor design to different values of defined parameters.

Only aircraft with certain equipage level are capable of flying in corridors. As a result, once a corridor is activated, all the aircraft without the necessary equipments must be rerouted out of corridors. Future work may analyze the additional system-wide cost due to this re-routing.

Our analysis was based on the assumption that system is quite predictable up to a timeframe T_p , during which there is no uncertainty in aircraft intent. Future work may consider data beyond T_p with lower validity weight.

Corridors are really the evolution of existing RNAV Q-routes. The existing Q-routes are published based on airline requests without an objective methodology for identifying their optimal location. The methodologies presented herein may also be used to calculate the optimal configuration of Q-routes.

ACKNOWLEDGMENT

This research was sponsored by NextGen Airspace Program at NASA Ames Research Center under order number NNA08BA50D. Authors wish to thank Mrs. Shannon Zelinski at NASA Ames Research Center for her support and contribution to this research.

REFERENCES

- [1] K. Corker, B. Gore, K. Flemming, and J. Lane, "Free flight and the context of control: Experiments and modeling to determine the impact of distributed air-ground air traffic management on safety and procedures," in *3rd USA/Europe Air Traffic Management R & D Seminar, ATM-2000*, Napoli, Italy, June 2000.
- [2] P. Kopardekar, N. Smith, K. Lee, A. Aweiss, P. Lee, T. Prevot, J. Mercer, J. Homola, and M. Mainini, "Feasibility of mixed equipage operations in the same airspace," in *8th USA/Europe Air Traffic Management Research and Development Seminar*, Napa, California, Jun. 2009.
- [3] J. Hoekstra, R. Ruigrok, R. van Gent, J. Visser, B. Gijsbers, M. V. Clari, W. Heesbeen, B. Hilburn, J. Groeneweg, and F. Bussink, "Overview of NLR free flight project 1997-1999," NLR, Tech. Rep. NLR-TP-2000-227, 2000.
- [4] P. Kopardekar, K. Bilimoria, and B. Sridhar, "Airspace configuration concepts for next generation air transportation," *Air Traffic Control Quarterly*, vol. 16, no. 4, 2008.
- [5] "Concept of operations for the next generation air transportation system," Joint Planning and Development Office, Washington, DC, Tech. Rep., June 2007.
- [6] A. Yousefi, G. Donohue, and L. Sherry, "High volume tube shaped sectors (HTS): A network of high-capacity ribbons connecting congested city pairs," in *23rd Digital Avionics Systems Conference (DASC)*. Salt Lake City, UT: IEEE/AIAA, 2004.
- [7] K. Sheth, T. Islam, and P. Kopardekar, "Analysis of airspace tube structures," in *27th Digital Avionics System Conference (DASC)*. IEEE/AIAA, Oct. 2008.
- [8] B. Sridhar, S. Grabbe, K. Sheth, and K. Bilimoria, "Initial study of tube networks for flexible airspace utilization," in *AIAA Guidance, Navigation, and Control Conference and Exhibit, Keystone, Colorado*. AIAA-2006-6768, Aug. 2006.
- [9] M. Xue and P. Kopardekar, "High-capacity tube network design using the hough transform," in *Proceeding of AIAA Guidance, Navigation, and Control Conference*. Honolulu, Hawaii: AIAA, Aug. 2008.
- [10] R. Hoffman and J. Prete, "Principles of airspace tube design for dynamic airspace configuration," in *AIAA-ATIO Conference Anchorage*. Alaska, USA: AIAA-2006-6768, Aug. 2008.
- [11] J. Alipio, P. Castro, H. Kaing, N. Shahid, O. Sherzai, G. Donohue, and K. Grundmann, "Dynamic airspace super sectors (dass) as high-density highways in the sky for a new us air traffic management system," in *Systems and Information Engineering Design Symposium*. Salt Lake City, UT: IEEE, April 2003.
- [12] D. Wing, J. C. Smith, and M. G. Ballin, "Analysis of a dynamic multi-track airway concept for air traffic management," NASA Langley Research Center, Hampton, VA, NASA/TP-2008-215323, Jul. 2008.
- [13] H. Hering, "Air traffic freeway system for europe," Eurocontrol Experimental Centre, Cedex, France, Tech rep. 20/05, November 2005.
- [14] A. Yousefi, J. Lard, B. Sliney, J. Timmerman, and F. Foreman, "Human-In-The-Loop experiments for initial evaluation of corridors-in-the-sky," Metron Aviation Inc., Dulles, VA, Tech. Rep. to NASA Ames Research Center NNA08BA50D, June 2009.
- [15] A. Yousefi, K. Stefanidis, M. Lowther, and A. Zadeh, "Airspace class scheduler: Analysis of aircraft equipage and traffic characteristics for determining optimal allocation of airspace classes," Metron Aviation Inc., Dulles, VA, Tech. Rep. NNA08BA50D, May 2009.
- [16] M. R. Jardin, "Real-time conflict-free trajectory optimization," in *5th USA/Europe ATM 2003 R&D Seminar*, Budapest, Hungary, June 2003.
- [17] "High altitude airspace redesign phase 1," U.S. Department of Transportation Federal Aviation Administration, Washington, DC, Advisory Circular 90-99, September 2003.
- [18] T. Myers, C. Bittle, D. Larsen, Y. Lee, and D. Rojas, "Nas simulation queuing model for evaluating proposed tmis," Metron Aviation Inc., Dulles, VA, Tech. Rep. submitted to FAA ATO 32F806-1209-008, 2009.

Coordinating multiple traffic management initiatives with integer optimization

Andrew M. Churchill, David J. Lovell

Department of Civil and Environmental Engineering &
Institute for Systems Research
University of Maryland
College Park, Maryland U.S.A.
churchil@umd.edu, lovell@umd.edu

Michael O. Ball

R.H. Smith School of Business &
Institute for Systems Research
University of Maryland
College Park, MD U.S.A.
mball@rhsmitr.umd.edu

Abstract—In this paper, the problem of coordinated resource rationing is considered in the face of current operational practice. In the United States, access to congested aviation resources is typically controlled by a system of capacity rationing wherein flights are assigned to slots at specific times. This is a well-accepted and efficient system, but it is not well-equipped to handle the problem faced when a single flight is controlled by more than one rationing initiative. The question of which, if any, initiative takes precedence over the others is not easily answered. An integer optimization model is introduced in this paper to find the delay-minimizing combination of multiple slot assignments for a set of flights and rationing initiatives. Rather than approach the problem comprehensively, this model treats each rationing initiative as somewhat independent, including only a constraint to guarantee that whatever slot pairs are assigned are mutually compatible. Computational results, including a case study, are reported along with directions for continuing research.

Keywords— ground delay program, airspace flow program, resource rationing, integer programming

1. INTRODUCTION

In air transportation, demand for certain resources occasionally meets or exceeds the available capacity. This situation may be particularly acute at certain airports for arrival and departure operations, and in the airspace along certain routes. At some of these resources, congestion may occur under even nominal conditions, but at others, weather is typically the exacerbating factor.

Several systems have been developed around the world to address these demand-capacity imbalances. The underlying principle in each of these systems is that demand-capacity imbalances should be addressed before the affected flights depart, as absorbing delay on the ground is safer and less expensive than doing so in the air. As a result, flights are typically given controlled departure or arrival times corresponding to some slots they have been assigned in the rationing initiative.

In the U.S., ground delay programs (GDP) and airspace flow programs (AFP) address demand-capacity imbalances expected at airports and in the airspace, respectively. These initiatives operate independently from one another and are employed sparingly – under nominal conditions they are

typically not used. In Europe, the Central Flow Management Unit (CFMU) assigns control times comprehensively to address demand-capacity imbalances throughout each flight's route. This system is employed continuously, always adjusting flight control times.

The process of assigning ground delays to alleviate these demand-capacity imbalances has been well-studied scientifically, having been first systematically outlined in [1]. The single initiative case, similar in principle to the GDP and AFP used in the U.S., has been formulated for deterministic [2] and stochastic ([3], [4]) cases, as well as with static [2] and dynamic ([4]) decision making structures. Likewise, the problem has been expanded to consider a network of airports over which individual aircraft operate multiple subsequent flights ([5], [6]). In all of these cases, however, each flight has only been affected by a single rationing initiative because only airport arrival capacity was rationed, and a single flight may only arrive once. Until the summer of 2006, airspace capacity was not explicitly rationed [7]. At this time, AFP was introduced, employing the same principles and software to manage disruptions as are used for GDP [8].

As ground holding strategies have been implemented in the U.S., there has been a strong desire to include user (e.g., airlines, private jets) priorities. This led to the development of the Collaborative Decision Making (CDM) paradigm [9]. This community established that the most equitable means by which capacity is rationed is the published flight schedule, although other metrics have been considered [10]. Through the user input this process allowed, numerous enhancements have been implemented, including exempting certain flights, facilitating slot trades, and crediting flight operators for providing timely information about flight status.

At the other end of the complexity spectrum from single initiative models used in practice are the models that assign delays comprehensively, while considering any and all resources a flight encounters along its route. Models such as this ([11], [12], [13], [14]) are more consistent with the European approach to managing demand-capacity imbalances, but face difficulties due to the tremendous data

requirements and complexity resulting from their comprehensive viewpoint.

An important distinction with these comprehensive models versus those described previously is that these must consider all flights simultaneously to build a plan, whereas the previous models consider only those flights explicitly affected by demand-capacity imbalances. While it is likely that such comprehensive approaches will minimize total delays, they have not been implemented in practice for a wide range of possible reasons. The single initiative models are more easily embedded into decision support system where human intervention is possible, especially to account for dynamically changing conditions. Related to this is the challenge of integrating the global models within the CDM paradigm. Our goal with the research presented here is to develop models that lie somewhere in between so as to achieve a more global perspective while preserving the important CDM and practical decision support features.

We seek to model an application setting where some flights are affected by multiple rationing initiatives, but each of these initiatives functions, to a large degree, independently. This is a less complex problem than the comprehensive approach, since it does not require planning for flights not affected by constrained resources. Considering these rationing initiatives in the independent fashion as is done today, with an eye toward how they may better be coordinated is the objective of this paper. A numerical example of this conflicting situation is provided as a further introduction to this problem.

To address this problem area, several models are shown. First, a formulation for the single resource rationing problem is defined. Then, using that as a basis, a formulation is shown for rationing several connected resources to ensure feasible slot pair assignments. Finally, computational results on a realistic case study demonstrating both the applicability and feasibility of this coordinated formulation are shown.

2. MOTIVATING EXAMPLE

The models presented in this paper address the U.S. paradigm of rationing capacity independently at each resource. Specifically, the problem of coordinating the potentially conflicting times assigned by multiple independent rationing initiatives is examined. The impact of this problem is examined in the following example illustrated by the space-time diagram shown in Figure 1.

In this example, a single flight is travelling from its origin airport to its destination and is passing through a stormy region of the airspace with reduced capacity. Further, capacity at its destination is reduced. Thus, it is subject to rationing both en route and at its destination. The nominal interoperation times at each resource are two minutes, but under the degraded conditions, these increase to three minutes. Rationing is performed according to the accepted principle of using the schedule as a baseline.

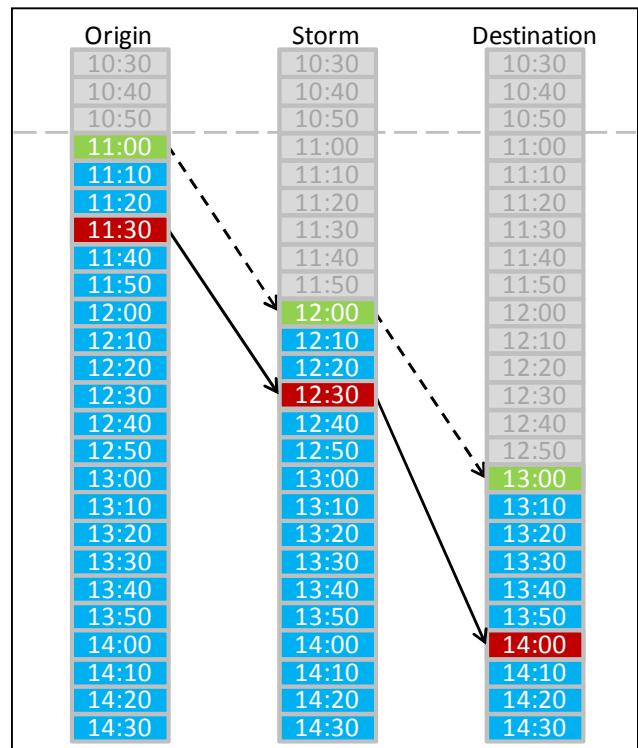


Figure 1 – Example of conflicting rationing

The total travel time is one hour, and the storm lies directly in the middle of the trip, as represented in the figure. The dashed lines represent the two hour scheduled trip departing at 11:00 that arrives at 13:00. The slot assignments after rationing are shown with the solid lines.

The first leg of the trip shown (origin-storm) does not present any problems with the assigned slots, as the flight can depart the origin at any time. However, using this slot at the storm will require a 90 minute travel time between the storm and destination. This is unacceptably more than the nominal travel time, resulting in an infeasible combination.

Clearly, however, this situation can be resolved by prioritizing one initiative over the other and exempting the flight through the secondary initiative by automatically granting it a slot other than the one it would have received. In the case of multiple flights affected by more than two different resources, the solution becomes much more difficult. The various considerations yield many possible combinations of outcomes. This is a complex combinatorial problem, and several approaches which leverage various properties of this problem are presented in the next section.

3. SINGLE INITIATIVE MODEL

Because the objective of this paper is to describe a model which coordinates multiple quasi-independent rationing initiatives, it is useful to first develop the model that represents the independent rationing process.

The broad objective of such a model is to ration capacity at a single congested resource both efficiently and equitably. The formulation shown here accomplishes this by assigning

flights to slots in a delay-minimizing fashion. While an integer optimization approach is employed, the results can be mimicked through algorithmic means. The model shown here is similar in principle to many others but is included as an important transition to the multiple initiative model in the subsequent section.

One feature that does make this formulation unique from previous research is the paradigm used to describe the rationed capacity. In this model, the index of the slot to which a flight is assigned is decoupled from the time associated with that slot. This adds an additional qualification (discussed later) to some of the summation terms, but in so doing, helps to reduce formulation size. This stands in contrast to the construct used by most ground holding models of constant length time intervals, each with varying capacity.

Rationing models in other research have been formulated with a uniform lattice of time periods over which discrete capacity changes occur. Typically these periods encompass more than one flight. Because airport capacities are specified as hourly rates, this approximation presents problems, primarily that the number of different hourly rates that can be represented in each time period is limited by the length of that time period. The length of these discrete periods is on the order of 5-15 minutes. For example, using a lattice with 5 minute bins enables each bin to represent hourly capacities that are multiples of 12.

In the formulations shown in the paper, capacity divisibility is enabled by using slots of unit capacity. Slots are not necessarily assigned on a uniform lattice, thus allowing for non constant interoperation times. Thus, resource capacity is specified as a list of slots, rather than a list of time periods each with associated capacity.

Because the models shown in this paper address deterministic capacity, the primary advantage of this modeling paradigm is the capacity divisibility. However, for stochastic formulations that consider specific potential outcomes, this methodology may greatly simplify formulations and solutions. For example, in cases in which capacity may only increase over time in various scenarios, this paradigm greatly simplifies solutions, as flights will stay assigned to the same slot index, but the operation time associated with that index will decrease.

A. Formulation

The model described here is an assignment formulation with limitations placed on the set of slots to which a given flight may be assigned. Several input data are required to understand this formulation.

The set F comprises the individual flights affected by the rationing initiative, each with prespecified scheduled arrival time α_f . The set S comprises the slots to which those flights will be assigned, each with associated slot beginning time τ_s . In this work, the number of slots always equals or exceeds the number of flights, as shown in (1). In this work, a

feasible solution is assumed to exist – this also implies the condition shown in (1). In building a case study, this is a trivial condition to enforce, as many additional slots with small interoperation times may be created after the planned initiative end time. This simulates the reality at most airports, at which operations may be extended late into the night to accept the day's flights.

$$|S| \geq |F| \quad (1)$$

The decision variables x_{fs} are integer valued, assuming a value of one when flight f is assigned to slot s and zero in all other cases. Decision variables are only created for combinations of f and s for which the condition in (2) is met. This helps to reduce the size of the constraint matrix by eliminating unnecessary variables. In principle, others could be considered, but would be necessarily fixed to zero.

$$\tau_s \geq \alpha_f \quad (2)$$

The first constraint set in this formulation is shown in (3). This enforces the condition that each flight must be assigned to exactly one slot. An additional condition is imposed that the slot to which each flight is assigned must begin at or after the flights scheduled arrival, as in (2).

$$\sum_{\substack{s \in S: \\ \tau_s \geq \alpha_f}} x_{fs} = 1 \quad \forall f \in F \quad (3)$$

The second constraint set, (4), enforces the capacity of each slot to be at most one flight. As discussed, the construct of using single-flight slots is also somewhat unique. Other models have assumed longer slot lengths to avoid the problem of capacity divisibility, but have then assigned multiple flights to each slot.

$$\sum_{f \in F} x_{fs} \leq 1 \quad \forall s \in S \quad (4)$$

The objective of this optimization problem is specified by the function shown in (5). This function minimizes the total sum of ground delays assigned to all flights. The superlinear function of delay length is used to favor the assignment of two short delays over a single long one. This principle contributes to equity between different flight operators because flights that are similar a priori are assigned similar delays.

$$\min z = \sum_{f \in F} \sum_{\substack{s \in S: \\ \tau_s \geq \alpha_f}} (\tau_s - \alpha_f)^{1+\epsilon} x_{fs} \quad (5)$$

B. Problem size

One measure of formulation strength and computational tractability is the size of the constraint matrix. The theoretical maximum/minimum numbers of constraints and variables (vary depending on slot spacing and flight characteristics) is shown in Table 1. Several realistic numerical problem sizes are shown as well.

TABLE 1 – SINGLE INITIATIVE PROBLEM SIZE

	Constraints	Variables	
		<i>Maximum</i>	<i>Minimum</i>
Nominal	$ F + S $	$\frac{ F }{2}(F +1)$	$ F S $
Small $ F =10$, $ S =10$	20	55	100
Typical $ F =100$, $ S =150$	250	5050	15000

4. MULTIPLE INITIATIVE MODEL

Although the first model shown in this paper has some interesting properties, the reality is that it addresses a fairly well-solved problem. In particular, its solution is also attainable through the use of the ration-by-schedule algorithm, given several assumptions. Of greater interest, however, is the case in which multiple rationing initiatives assigning conflicting slot times to flights. This problem is addressed in this section.

The essentials of this integer programming formulation are similar to those shown in the previous section, with the exception of the addition of a single constraint set. This is, of course, the stated objective of this work – to develop a coordinated rationing method consistent with current practice. The added constraints enforce the logical condition that the slot times assigned to flight that use multiple initiatives are compatible. In some ways, this model is an extension and simplification of the concept proposed in [15], in that only those regions under adverse conditions are expressly controlled. However, the application considered is more specific than the system-wide plan developed in that work and the modeling approach completely different.

Thus, little additional input data are required. The set I comprises the initiatives that are to be rationed. As a result of this addition, each initiative has its own independent slot set that is indexed with i as S^i . In addition, the slot times and scheduled flight arrival times are each now indexed by i as well. If the number of initiatives is one, then this formulation simply reduces to that shown previously. Thus, this model may be seen as a generalization of the previous.

Several input data are required to account for the multiple initiatives for each flight. The set V_f is defined as all initiatives visited by flight f . For a flight to be included in this model, $|V_f|$ must be greater than zero. The value N_f^i is the initiative visited by flight f after initiative i . This is used to maintain the ordering of initiatives.

A. Formulation

The decision variables x_{fs}^i are integer valued, assuming a value of one when flight f is assigned to slot s in initiative i and zero in all other cases. Decision variables are only created for combinations of f , s , and i for which the condition in (6) is met. Similar to condition (2), this helps to reduce the size of the constraint matrix by eliminating unnecessary variables.

$$\tau_s^i \geq \alpha_f^i \quad (6)$$

The first constraint set in the linked formulation is shown in (7). It is similar to that shown in (3), with the added dimension of each initiative. This enforces the condition that each flight must be assigned to exactly one slot in each rationing initiative.

$$\sum_{\substack{s \in S^i : \\ \tau_s^i \geq \alpha_f^i}} x_{fs}^i = 1 \quad \forall f \in F, i \in I : i \in V_f \quad (7)$$

Likewise, constraint set (8) is similar to (4) in that it enforces the condition that each slot in each initiative may have at most one flight assigned to use it.

$$\sum_{\substack{f \in F : \\ i \in V_f}} x_{fs}^i \leq 1 \quad \forall i \in I, s \in S^i \quad (8)$$

The constraint set that links together these multiple initiatives is shown in (9). This constraint set works by defining feasible slot combinations in each pair of initiatives for each flight. The range R_{fs}^{ij} defines the times that flight f could feasibly arrive at initiative j by using slot s in initiative i . For each flight f , some maximum delay between initiatives m_f is defined.

$$\begin{aligned} & \forall f \in F, i \in V_f, \\ & x_{fs}^i - \sum_{\substack{t \in S^j : \\ \tau_t^j \in R_{fs}^{ij}}} x_{ft}^j \leq 0 \quad j = N_f^i, s \in S^i : \\ & \tau_s^i \geq \alpha_f^i, N_f^i \neq \emptyset \end{aligned} \quad (9)$$

The maximum inter-initiative delay value should be small, as it defines the period over which the operator is indifferent to various slot assignments. That is, the unit cost of these first few minutes of “airborne” delay is assumed to be equal to the unit cost of ground delay. The maximum delay in this context is meant only to permit a bit of slack for the variations in spacing of slot times at each initiative.

The concept of constraint set (9) is illustrated in Figure 2. Assume that there are two initiatives (i and j), and the nominal travel time between them is 60 minutes. Further, assume that a flight may travel somewhat slower than nominal between the two, up to six extra minutes. This six extra minutes then forms the maximum inter-initiative delay value. Under this scenario, the feasible slots to use in the second initiative are shown for the 12:06 slot in the first.

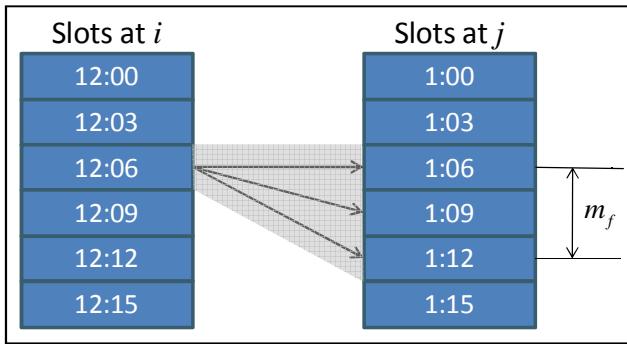


Figure 2 – Feasible range example

The range R_{fs}^{ij} used in (9) is defined in (10) as beginning at the sum of the time for slot s at initiative i and the inter-initiative travel time. The range ends after the maximum inter-initiative delay.

$$R_{fs}^{ij} = [\tau_s^i + \alpha_f^j - \alpha_f^i, \tau_s^i + \alpha_f^j - \alpha_f^i + m_f] \quad (10)$$

The objective of this formulation, as with other models considering air traffic management problems, is to minimize delays. It is very important, however, to consider precisely which delay is being minimized. Two alternate objectives, considering different scopes of delay, are presented here.

The difference between the two possible objectives is the scope of the delay summed. The first considers delays at each initiative independently while the second considers only those at a flight's final initiative. This first potential objective, considering the total amount of delay assigned, is shown in (11). Thus, some delays may be "double counted" according to (11) because they are counted twice but truly impact the flight only upon arrival to its destination.

$$\min z = \sum_{f \in F} \sum_{i \in I} \sum_{s \in S: \tau_s^i \geq \alpha_f^i} (\tau_s^i - \alpha_f^i)^{1+\varepsilon} x_{fs}^i \quad (11)$$

The second objective, considering only the delay at the flight's final initiative, is shown in (12).

$$\min z = \sum_{f \in F} \sum_{i \in I} \sum_{s \in S: \tau_s^i \geq \alpha_f^i \text{ and } N_f^i = \emptyset} (\tau_s^i - \alpha_f^i)^{1+\varepsilon} x_{fs}^i \quad (12)$$

Because this model treats capacity deterministically and sufficient capacity at the origin airport is assumed, all delays are taken on the ground before departure. Thus, there is no need to consider a cost differential. A comprehensive model that schedules all points along a flights route, or one that considers capacity stochastically, however, would fail for this assumption.

B. Formulation size

Again, the size of the constraint matrix is considered. The theoretical worst case numbers of constraints and

variables and several realistic numerical problems are shown in Table 2.

TABLE 2 – LINKED INITIATIVES PROBLEM SIZE

	Constraints	Variables
Nominal	$\sum_{f \in F} v_f + \sum_{i \in I} S^i + \sum_{f \in F} v_f * \sum_{i \in I} S^i $	$ F \sum_{i \in I} S^i $
Small $ F = 10, I = 2, S^i = 10$	440	200
Typical $ F = 100, I = 3, S = 150$	135750	45000

5. COMPUTATIONAL RESULTS

To demonstrate the efficacy of this linked rationing initiative formulation, a realistic numerical case study has been developed. The schedule data are randomly generated, but represent a realistic situation such as is encountered during summer convective weather over the northeastern United States. Both objective functions described are considered, as is a heuristic mimicking current practice.

In this case study, there are two airports (B, C) for which capacity is being rationed, and one disruption (A) in the enroute airspace for which rationing must take place, as shown in Figure 3.

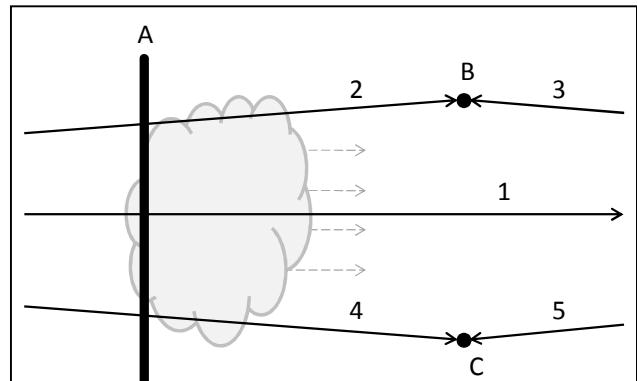


Figure 3 – Case study layout

To better visualize the problem setup, the results are discussed in terms of flows, as labeled in this figure. Flow 1 comprises flights crossing the enroute disruption, but not travelling to either of the two disrupted airports. Flows 3 and 5 travel to airports B and C, respectively, but do not cross the enroute disruption. The most interesting flows, 2 and 4, cross the disrupted airspace before arriving at the disrupted airports (B and C, respectively). It is these two

groups of flights that confound the traditional single resource ration-by-schedule methods employed.

Nominally, each resource has a capacity of 60 flights/hour (or an interoperation time of 60 seconds), but under the reduced conditions in this case study will have half of that capacity available for a four hour period. The number of scheduled arrivals to each resource over the study period is shown in Figure 4. Because the travel time between the storm and each airport is one hour, the bars representing the schedule for flows 2 and 4 are simply shifted by one hour from their appearance in the resource A schedule to their appearances in the resource B and C schedules, respectively.

The schedule is assumed to terminate after the flights shown in Figure 4. While potentially unrealistic, this simplifies considerably the conditions surrounding the end of the program because the flights expected to arrive after the end of the program are not subject to rationing.

Because the capacity reduction is sufficiently extreme relative to the scheduled number of aircraft, the optimization model will assign flights to every slot. Thus, the time-varying profile of flights after the model has run will match precisely with the reduced capacity line until the entirety of the set of flights has been assigned.

This case study resulted in a constraint matrix with 24540 rows (constraints) and 189992 columns (variables). The model was run on a quad processor system with 16GB memory using the Xpress 2008A solver. Two objectives were considered. Using (11), the model solved to integer optimality with the linear programming (LP) relaxation in 38 seconds. With (12), the LP relaxation was not integer and so branch and bound was employed. After 60 minutes, the model had a 13.0% optimality gap, while after 240 minutes, this had narrowed to 2.5%. The obvious differences resulting from these two objectives warrant further investigation.

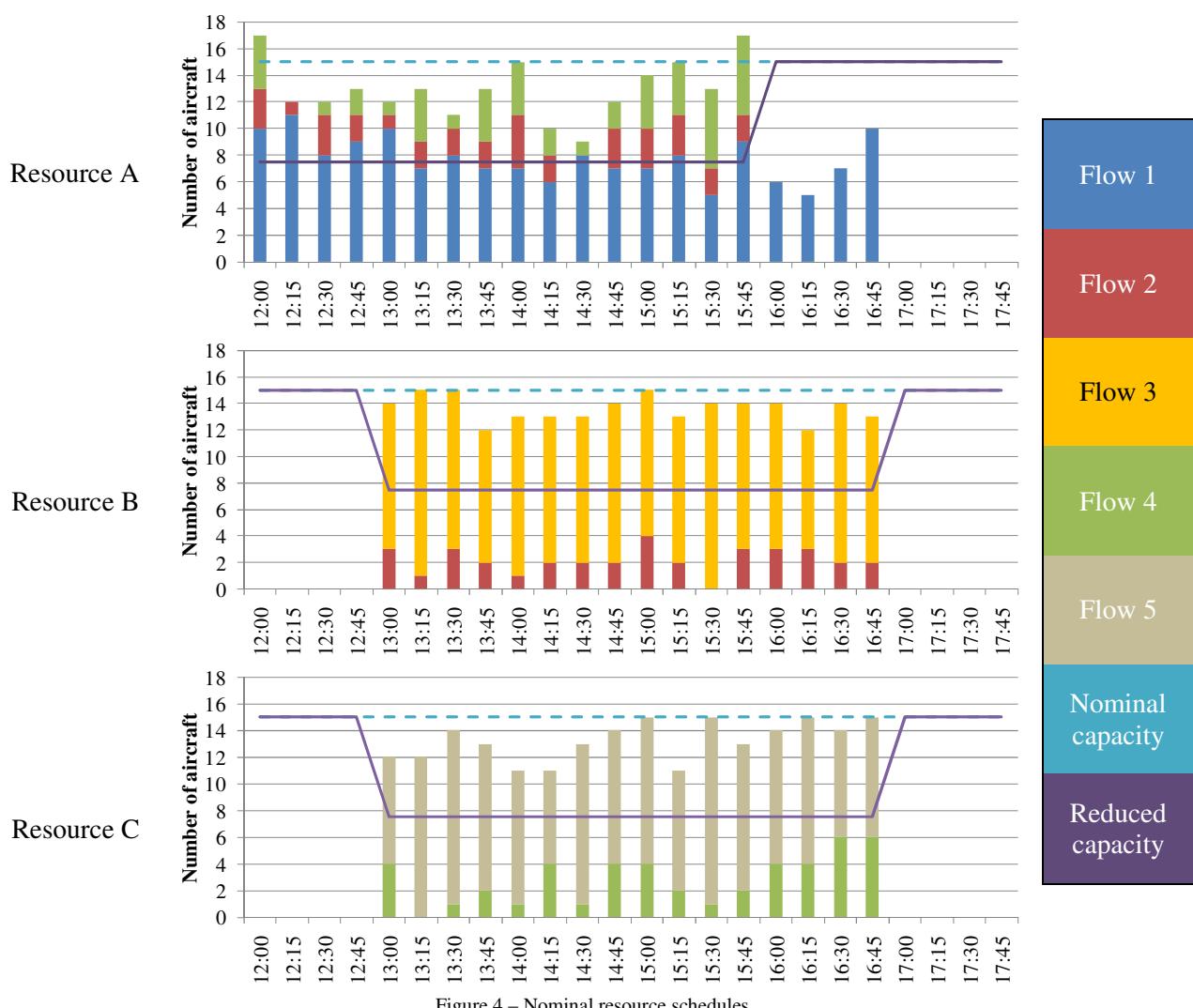


Figure 4 – Nominal resource schedules

The results for the two objectives and one heuristic approach are shown in Table 3. These results will be discussed in response to three issues regarding delay distribution:

1. Within the total delay optimal results, but between the five flows: Does the model favor one flow or type of flow?
2. Between total delay and final initiative optimal results: How do these objectives affect each flow?
3. Between the total delay optimal results and a priority scheme mimicking current practice: How is the optimal solution different from the heuristic one?

The heuristic discussed in this section is a simple greedy one. First, all airport capacity is allocated, with flights also using the airspace initiative automatically given priority to proceed unimpeded. After those flights have been assigned, then the remaining airspace capacity is allocated to other flights according to schedule priority. This approach is consistent with that taken in practice today.

A. Issue 1: Fairness between flows

Using the total delay objective, the first issue examined is whether the model favors one flow or type of flow over another. The concern in this case is that flights using multiple initiatives may be unduly prioritized or penalized. A *t*-test at 5% significance suggests that the average delay assigned to each flow is not significantly different (at a 5% level) than the overall average delay for flows 1-4. The null hypothesis of average delay equal to overall average delay cannot be rejected for flow 5 however. There is no obvious structural explanation for this, so further investigation with other case studies is warranted.

TABLE 3 – SUMMARY OF AVERAGE DELAYS (IN MINUTES) ASSIGNED TO EACH FLOW (STANDARD DEVIATIONS SHOWN IN PARENTHESES)

Flow	Flight count	Total delay optimal	Final init. optimal (2.5% gap)	Heuristic
1	155	66.9 (26.9)	13.4 (7.3)	64.7 (25.5)
2	70	74.3 (31.3)	194.4 (62.7)	78.0 (33.9)
3	183	76.5 (34.1)	55.7 (26.9)	76.4 (33.9)
4	92	74.9 (26.8)	161.7 (79.3)	74.1 (28.6)
5	166	65.3 (31.9)	43.0 (22.5)	65.8 (31.8)

B. Issue 2: Differences between objectives

To examine the differences between the two objective functions considered, the average delays assigned to each flow are compared pair-wise. Clearly these results are different, particularly for the delays assigned to the flights in flows 2 and 4 – those that use two resources. A two sample

t-test at 5% confirms that each of the average delays assigned to these pairs of flows are significantly different. Although it is clear that the solutions are different, the reasoning behind these differences is not immediately clear and further investigation is warranted on this issue, particularly in light of the differences in solution times and the system-level attractiveness of the second objective function.

C. Issue 3: Comparison to heuristic

To examine the utility of the optimization approach with respect to solution times, a heuristic solution is also considered. This heuristic mimics current practice in that the schedule is used as the basis for rationing and conflicts between initiatives are resolved by prioritizing the airport initiative.

Comparing the heuristic and the total delay objective with a two sample *t*-test at 5% significance, the null hypothesis of equal means cannot be rejected for any matched pair of flows. This suggests that the distribution of delays across flows assigned by the optimization model is not statistically different from that assigned by the heuristic.

This result is promising for several reasons. Because the optimization models may take considerable time to solve depending on the objective, number of flights and slots, and other considerations, it is useful to have heuristics that can ably mimic their results. Thus, in a laboratory environment, the optimization models can be used as the standard by which heuristics may be judged. Further, the similarities in results for this case provide evidence that the approach employed for this problem in practice today is actually fairly efficient.

6. CONCLUSIONS

In this paper, a model that coordinates multiple independent resource rationing initiatives employed in air transportation was shown. The problem that this model solves lies between other air traffic management models currently considered in the literature, both with respect to complexity and scope. An optimization-based model was defined, its computational properties explored, and a realistic case study outlining its utility shown.

The results of the case study suggest that further exploration is needed to understand the different possible objectives employed in optimizing the distribution of delay. Clearly the solution resulting from the final initiative objective would be unacceptable in practice because of the tremendous inequity placed upon a subset of flights. The initial results of the case study, however, suggest that it is possible to develop a heuristic approach that mimics the solutions derived from the optimization model.

Several enhancements are possible to extend both the realism and utility of this formulation. The first such extension is the inclusion of additional input data and slightly modified constraints to allow for the consideration

of flight exemptions, both for flights already en route, or for those exempted for other reasons. In addition, more sophisticated methods for examining problem feasibility may be included with the addition of a single slot at a time much later than the last scheduled flight to capture flights that cannot otherwise find a feasible assignment.

In addition, the nature of the two different objective functions must be explored, aiming toward explaining what structural difference resulted in such a wide variance in solution times and delay distribution. Further, it may be interesting to explore other objective functions that make the trade between equity and efficiency differently from those shown here.

Finally, considerable attention will be paid to developing a formulation of this model that incorporates stochastic recourse. Of course this is a considerably more complex problem that departs from practice, but previous research on including stochastic information in air traffic management problems has shown that great potential for insight and delay savings exists.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support of the National Aeronautics and Space Administration Airspace Systems Program under ARMD NRA: NNH06ZNH001 and Metron Aviation for assistance with data.

REFERENCES

- [1] Odoni, A.R., "The flow management problem in air traffic control," in *Flow Control of Congested Networks*, Amedeo R Odoni and G Szego, Eds. Berlin: Springer-Verlag, 1987.
- [2] Terrab, M. and A.R. Odoni, "Strategic flow management for air traffic control," *Operations Research*, vol. 41, no. 1, pp. 138-152, 1993.
- [3] Ball, M.O., R.L. Hoffman, A.R. Odoni, and R. Rifkin, "A stochastic integer program with dual network structure and its application to the ground-holding problem," *Operations Research*, vol. 51, no. 1, pp. 167-171, 2003.
- [4] Mukherjee, A. and M. Hansen, "A dynamic stochastic model for the single airport ground holding problem," *Transportation Science*, vol. 41, no. 4, pp. 444-456, 2007.
- [5] Vranas, P.B., D.J. Bertsimas, and A.R. Odoni, "The multi-airport ground-holding problem in air traffic control," *Operations Research*, vol. 42, pp. 249-261, 1994.
- [6] Vranas, P.B., D.J. Bertsimas, and A.R. Odoni, "Dynamic ground-holding policies for a network of airports," *Transportation Science*, vol. 28, no. 4, pp. 275-291, 1994.
- [7] Brennan, M., "Airspace flow programs - a fast path to deployment," *Journal of Air Traffic Control*, vol. 49, no. 1, pp. 51-55, 2007.
- [8] Krozel, J., R. Jakobovits, and S. Penny, "An algorithmic approach for airspace flow programs," *Air Traffic Control Quarterly*, vol. 14, no. 3, pp. 203-230, 2006.
- [9] Wambsganss, M., "Collaborative decision making through dynamic information transfer," *Air Traffic Control Quarterly*, vol. 4, no. 2, pp. 109-125, 1997.
- [10] Ball, M.O., R.L. Hoffman, and A. Mukherjee, "Ground delay program planning under uncertainty based on the ration-by-distance principle," *Transportation Science*, 2009.
- [11] Bertsimas, D.J. and S. Stock Patterson, "The air traffic flow management problem with enroute capacities," *Operations Research*, vol. 46, no. 3, pp. 406-422, 1998.
- [12] Bertsimas, D.J. and S. Stock Patterson, "The traffic flow management rerouting problem in air traffic control: a dynamic network flow approach," *Transportation Science*, vol. 34, no. 3, pp. 239-255, 2000.
- [13] Lulli, G. and A.R. Odoni, "The european air traffic flow management problem," *Transportation Science*, vol. 41, no. 4, pp. 431-443, 2007.
- [14] Bertsimas, D.J., G. Lulli, and A.R. Odoni, "The air traffic flow management problem: an integer optimization approach," in *13th International Conference, IPCO 2008*, Bertinoro, Italy, 2008, pp. 34-46.
- [15] Churchill, A.M., D.J. Lovell, and M.O. Ball, "Evaluating a new formulation for large-scale traffic flow management," in *Proceedings of the 8th USA/Europe Air Traffic Management R&D Seminar*, Napa, California, 2009.

Collaborative Rerouting in the Airspace Flow Program

A Framework for User-cost Based Performance Assessment

Amy M. Kim, Mark Hansen

Department of Civil & Environmental Engineering
University of California, Berkeley
Berkeley, CA
amy_kim@berkeley.edu

Abstract— The Airspace Flow Program (AFP) ground delays flights in order to control their flow through capacity constrained airspace regions. It has been successful in controlling traffic with reasonable delays, but the procedures must be improved upon to handle future projected demands. This paper explores a future AFP where centrally-managed rerouting and user input are incorporated into the initial resource allocations. A modeling framework was developed to evaluate and compare allocation strategies, under differing assumptions about traffic managers' knowledge about airline flight costs. It is used to quantify tradeoffs regarding the quality and timing of airlines' input information. Three allocation strategies were developed; they differ with respect to the input requested of airlines, and the resource allocation philosophy. They are assessed based on the total cost impact of the AFP initiative on flight operators. To this end, a flight cost function was developed to represent the cost of delay specific to each flight; it consists of deterministic components to represent what traffic managers know about the airlines, and a stochastic component to represent that which they do not. A numerical example demonstrates the situations under which better information quality could be more desirable than timeliness, and vice versa. Identifying these types of tradeoff points is a key contribution of this research effort.

Keywords- *delay; air traffic flow management (ATFM); Airspace Flow Program (AFP); Collaborative Decision Making (CDM); user cost; strategic planning.*

I. INTRODUCTION

Adverse weather frequently and severely impacts flight operations in the National Airspace System (NAS). In addition, with the growth in demand projected over the next 20 years, weather and traffic-induced delays are also anticipated to increase under the current system. Air traffic flow management (ATFM) programs are used to reduce the scale and cost of disruptions to flight operators. One such initiative is the Airspace Flow Program (AFP), in which flights are held on the ground at departure airports in order to meter them through capacity constrained airspace regions. The AFP was first implemented in 2006 in the northeast region of the U.S., and has proven to be successful in controlling traffic with reasonable flight delays. However, as demands increase into the future, the benefits derived from the AFP process will become limited unless a procedure to better utilize airspace capacity is incorporated into the process.

This research addresses the need for a more comprehensive, centrally-managed, and user-input based resource allocation program for AFPs. We develop a modeling framework through which we formulate, evaluate, and compare strategies that employ rerouting combined with ground delay to minimize the impacts of AFP initiatives on users of the NAS. The assignment strategies differ with respect to the inputs requested of users, and the rules by which resource allocation decisions are made. This paper presents three strategies based on combinations of two resource allocations schemes and two forms of user input. The main objective of this paper is to investigate how these strategies perform in comparison to one another under different assumptions about airline utility. The model framework through which we can identify the tradeoff points between strategies is a key contribution of this research.

Throughout this paper, “operator” will be used to refer to NAS users such as commercial airlines and general aviation aircraft. “Traffic manager” will refer to traffic managers overseen by the Federal Aviation Administration (FAA). Section II describes the current system, and a literature review. Section III contains a problem overview while Section IV introduces the modeling framework and models. Section V presents an illustrative numerical example and Section VI concludes with a discussion and plans for future work.

II. BACKGROUND

A. Constrained Airspace Rerouting

Flight rerouting due to severe en route weather and traffic congestion is performed in both strategic and tactical ATFM. It is manually intensive as it requires close coordination between several traffic management units. As a result, traffic managers select reroutes from a standard set compiled in the National Playbook, basically employing a “one size fits all” approach [1] without input from the operators. Airlines also have the option of rerouting their own flights before and after departure, subject to traffic managers' approvals. They often exercise this option to avoid assignment of undesirable routes and heavily delayed departure times.

Air traffic flow management initiatives, including centralized rerouting, can be inefficient without input from users, because resource allocations are made without knowledge about the value of the assignment to users. As a

This work is sponsored by the Federal Aviation Administration (FAA).

result, more collaborative approaches to rerouting have been proposed. Concepts that aim for more structured coordination between traffic managers and operators have existed since the early 2000s, but implementation has been difficult.

B. Collaborative Decision Making (CDM)

A significant improvement to NAS air traffic management began in the mid-1990s with the Collaborative Decision Making (CDM) program. CDM is a joint government and industry initiative that aims to improve both the technological and procedural aspects of air traffic management, through improved information exchange between government and industry. The first major application of CDM was to Ground Delay Programs (GDPs). When an airport has reduced arrival capacity due to severe weather either en route or near the airport, a GDP holds flights destined for that airport on the ground at their origin airports to meter demand. CDM information exchange between operators and the traffic manager drastically enhanced the effectiveness of GDPs in correcting demand/capacity imbalances and reducing delays, by ensuring that the traffic manager have up-to-date demand information and that "slots" vacated as a result of cancellations or other events could be used for other flights. GDPs are very effective in managing reduced arrival capacity when it is caused by inclement weather near the destination airport. However, GDPs can be inefficient, ineffective and inequitable in addressing en route constraints. As a result, the AFP was first implemented in 2006 to handle en route constraints.

C. Airspace Flow Program (AFP)

In an AFP, the constrained airspace region and the flights filed into this region during the time of reduced capacity are identified. The reduced capacity is then distributed by assigning delayed departure times to the impacted flights. Constrained airspace regions include those that are experiencing undesirable weather and/or heavy demands. Most AFPs begin after 2PM local time as airspace congestion and convective weather are more likely to occur after this time. They typically end after 10PM.

An AFP flight will receive a delayed departure time on its original filed route. It can either accept the assigned departure time, or reject and reroute around the constrained airspace (subject to traffic managers' approval). Slots to fly through the constrained region are vacated as flights are canceled or routed out, and the schedule is compressed such that remaining flights are moved up in time. Currently, the distribution of delayed departure times combined with airline-initiated rerouting and cancellation has proven to be adequate for handling capacity constraints. However, with growing demand, greater utilization of other available airspace capacity will be required. One strategy is to incorporate reroutes into the initial resource allocation, such that delayed departure times are combined with new route assignments. Flying a longer alternative route with less ground delay might be a desirable alternative to accepting a long ground delay on the original route¹. Also, if neighboring routes could be more optimally utilized, the total delay cost of the AFP could be reduced. In order to offer

¹ This would be particularly true when there are critical downline flight and crew connections to be made.

resource assignments that are desirable to operators, however, the FAA will require a significant level of user input.

D. Literature Review

There has been much work in developing optimization models to support ATFM decisions. The objective of many such models is to minimize the system-wide cost of delay. They consider ground holding, air holding and rerouting decisions. The Bertsimas and Stock-Patterson model provide for flight-specific air and ground hold cost ratios in their model, but do not provide any information about them [2]. Goodhart's models provide a framework where ATFM decisions are made through information exchange between the FAA and operators [3].

Uncertainties in weather and capacity have been addressed in the single airport ground holding problem, which has been considered in deterministic and stochastic, static and dynamic formulations. The earliest work began with [4] and [5], and the problem was addressed in a collaborative context by [6]. Reference [7] formulated an algorithm to schedule, reroute and airhold flights flying into and around constrained airspace, imposing ordering schemes that align with CDM.

Much literature exists about resource rationing and equity in ATFM, specifically within the context of GDPs. Reference [8] describes a framework for equitable allocation, illustrating their operational impacts and use in reducing systematic biases. Reference [9] compares the efficiency of airspace resource allocation schemes as alternatives to GDP allocation schemes.

The assumption of continuously distributed VOT over flight populations has not been studied in the context of the ATFM problem. Value of time (VOT) was examined as a continuous distribution [10] over a vehicle population for a steady-state congestion pricing model. Comparing different methods of incorporating heterogeneous users' preferences into ATFM models has also not been studied extensively.

III. PROBLEM OVERVIEW

The AFP facilitates resource allocation decisions when en route demand/capacity imbalances exist. In addition to system capacity constraints, under the CDM philosophy decisions are shaped by the allocation and equity principles chosen for implementation, as well as the user information provided to the process. By altering these inputs, the resulting allocation structure can potentially look very different from another.

There are many resource allocation schemes that could be considered [11] and we list a few. Traffic managers may be instructed to meet system cost targets with or without certain equity constraints. Users could be allocated resources by order of information submission, the original schedule, or a random order. Airlines could also be assigned a proportion of the total available resources based on the number of flights they have scheduled. Priority may be given based on aircraft size.

Performance assessments are based on system efficiency measures as well as user satisfaction and cost considerations, which are part of the users' utility structure. The overall performance of an allocation scheme will improve when inputs that well represent users' utility are incorporated. User inputs can come in many forms, and we introduce two in this paper.

Flight	Original Departure Times	Cost (b4 gr delay):		Allocation 1	Route/Slot	Cost
		Route 1	Route 2			
A	0	100	150		A: R1,S1	$100+(5-0)=105$
B	5	90	140		B: R1,S2	$90+(60-5)=145$
					Total	250

	Route 1	Route 2	Allocation 2	Route/Slot	Cost
Slot 1	5	0		A: R2,S1	$150+(0-0)=150$
Slot 2	60	20		B: R1,S1	$90+(5-5)=90$
				Total	240

Figure 1. Allocation Illustration

Consider the simple example illustrated in Figure 1. Two flights (A and B) are planned to travel some nominal route with original departure times 0 and 5 minutes. The route is closed due to convective weather; to accommodate these flights, departure slots on two alternative routes are offered. Say that flights A and B offered their en route costs (in ground delay minutes) for each route, shown in the top left. The final cost is calculated based on the difference between the original departure time and the new slot time, plus the en route cost. If traffic managers are obliged to serve Flight A first (Allocation 1), then Flight A would be given Route 1 slot 1 as it is the lowest cost option available to it. Flight B would be left with Route 1 slot 2 as its best available option. The total cost of this allocation is 250. If the goal is to minimize total cost (Allocation 2) they would assign Flight A to Route 2 slot 1 and Flight B to Route 1 slot 1. The cost of this allocation is 240. Clearly the allocation results could change if airlines submitted different cost values.

In this paper we consider a functional form to represent the cost of an AFP reroute to flights. This cost function has both deterministic and random components, to represent what information the FAA does and does not have about the operators of these flights. We use this function to assess the performance of several different resource allocation/user input combination models. We build models based on two different user input types – the parametric model and the stated route preference model. The parametric model requires users to supply parameters of the cost function, which traffic managers use to calculate costs of various reroute and ground delay options. The stated route preference model requires operators to supply more detailed, complete cost information about the route and ground delay options available in each AFP. It is based on the delay thresholds concept developed as part of the Flow Constrained Area Rerouting (FCAR) Decision Support Tool by Metron Aviation [12], which is discussed in further detail in the following section. Both models allocate resources based on system-optimal cost minimizations where equity is not considered. However, we also consider another version of the stated route preference model where flights are assigned resources by the order they submit their input data. In the first-submitted, first-assigned (FSFA) model, the earlier flights offer their input data, the more likely they are to receive a more desirable allocation. The FSFA allocation scheme is an easily understood and well-accepted rationale that has been adopted in various forms within CDM [9].

The main objectives of this research are to determine how models with these different resource allocation schemes and

user inputs perform against one another under changing assumptions about flight utility. Performance will be measured using the total generalized flight cost of each models' optimal AFP resource allocation. The result is a framework through which user input and resource allocation combinations can be represented, evaluated, and compared.

IV. MODEL FRAMEWORK

A. Evaluation Scenario

We introduce a simple model context in Figure 2. A nominal route (Route 1) connects two fixes in en route airspace. Flights enter Route 1 at entry fix “A” and leave at exit fix “B”. Route 1 has sufficient capacity to serve the scheduled demand $D_0(t)$, until a capacity constraint develops at a fixed location along its path and lasts for duration T . The capacity of Route 1 is reduced, and an FCA is created. The total scheduled demand must be reassigned to observe this reduced capacity. All N flights originally scheduled to use Route 1 are either given delayed departure times, rerouted to an alternative route, or both. Each alternate route r is characterized by its travel time air_r and capacity $S_r(t)$. We assume that fixes A and B are not bottlenecks, and for the purpose of this analysis they are considered the flights' origin and destination.

As mentioned previously, FAA traffic managers have limited access to the details of how airlines make flight cost calculations and subsequent routing decisions. This analysis is unconcerned with the airlines' actual costs for the original scheduled flight plans, as it is assumed that these flight plans were those most preferred under ideal conditions. We are concerned with evaluating the additional cost of greater en route time and ground delay due to AFP.

We can assume that $c_{n,r}$, the additional cost of the n^{th} departing flight taking route r due to an AFP, is a function of the increased air time (compared to the nominal route, and assuming that aircraft fly at fuel-efficient speeds) and time spent in ground delay. The additional en route time and ground delay account for many direct costs such as additional fuel, crew time, equipment maintenance, and indirect costs such as passenger satisfaction, gate time, flight coordination, and the airline's satisfaction with their own particular objectives. We assume air holding is not necessary because we have perfect information about the capacity constraint duration T , scheduled demand $D_0(t)$, and all route capacities $S_1(t), \dots, S_R(t)$. As such, all anticipated delay is incurred on the ground.

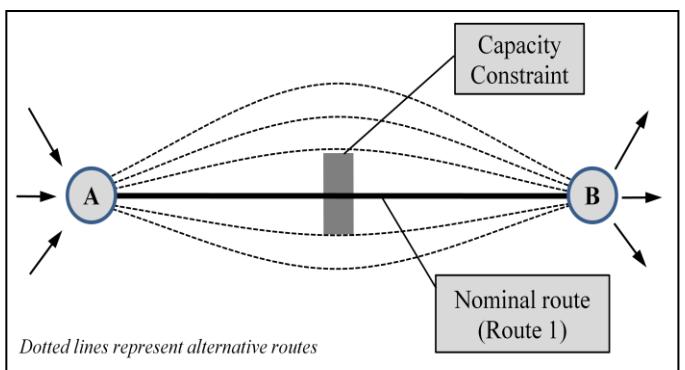


Figure 2. Model Framework

The generalized flight cost function is specified such that the air time, ground delay, and error components do not interact with one another. It is also a linear function of inputs, and is quantified in units of ground delay minutes.

$$c_{n,r} = c_{n,r}^{air} + c_{n,r}^{grdelay} + \varepsilon_{n,r}, \quad \varepsilon_{n,r} \sim P \quad (1)$$

Each cost component can be further identified as follows:

$$c_{n,r} = \alpha_n \cdot (h_r - h_0) + d_{n,r} - s_n + \varepsilon_{n,r}, \quad \varepsilon_{n,r} \sim P \quad (2)$$

where α_n is a ratio for converting additional AFP-related en route time to ground delay minute units for flight n , h_r is the newly assigned en route time for route r , h_0 is the en route time for the original (scheduled) route, $d_{n,r}$ is the new departure time for flight n on route r at fix A under the AFP, s_n is the original scheduled departure time for flight n at fix A, and $\varepsilon_{n,r}$ is the random error term for the cost of the AFP, and follows distribution P .

The quantity $(h_r - h_0)$ is non-negative because it is likely that the nominal route had the shortest flying time under an optimal speed, hence its status as the nominal route. Here we assume the effects of tactical control are insignificant compared to the delay cost of the AFP. Ground delay is non-negative because aircraft cannot depart before their original scheduled time, such that $(d_{n,r} - s_n) \geq 0$.

If the AFP capacity of each alternative route r is $S_r(t)$, it then follows that the instantaneous minimum headway at time t is $S_r^{-1}(t)$. Now assume that $S_r(t)$ is constant over the duration of the AFP, and aircraft on route r are scheduled with constant headways. We have established that the n^{th} flight (out of a total AFP population of N) is scheduled to depart at $d_{n,r}$. If we instead tabulate flights by route, the departure time of flight i on route r (of total flights X_r assigned to r) can be expressed as a linear function of i with slope $g_r = \frac{1}{S_r(t)}$. We also assume that original scheduled demand $D_0(t)$ is constant, and s_n can be expressed as a linear function of n with slope $g_0 = \frac{1}{D_0(t)}$. It then follows that the total estimated cost of an AFP (without accounting for unknown cost components) is expressed as:

$$\hat{C} = \sum_{r=1}^R \sum_{i=1}^{X_r} \alpha_i \rho_r + g_r i - g_{0,i,r} \quad (3)$$

where \hat{C} is the total estimated cost of an AFP, α_i is the cost ratio of additional AFP-related en route time for the i^{th} flight on route r , ρ_r is the additional en route time if a flight is reassigned to route r ($\rho_r = h_r - h_0$), g_r is the new AFP departure headway on route r , and $g_{0,i,r}$ is the original (before AFP) scheduled departure time for flight i on route r . Also, if X_r is the total number of flights assigned to route r , then $\sum_{r=1}^R X_r = N$.

This paper focuses on the case where all flights were originally scheduled to depart at the same time (i.e. $g_{0,i,r} = 0$). However, this analysis has been extended to a case where flights are originally scheduled to depart at different times.

B. Parametric Reroute Model

1) Concept

In the parametric reroute model, the FAA allocates AFP resources using the cost function shown previously (1-3) with parameters supplied by operators. If specified well, the model can provide a good reflection of operator utility, and the resource allocation can be very efficient. If specified poorly, resource allocations can be inefficient. We would like to ascertain how this approach performs in comparison to the stated route preference strategies under increasingly errant specifications.

We envision that an FAA mandate would require airlines to provide cost parameters for their domestic flights to a central database. Airlines would be encouraged to update these parameters as desired. When an AFP is announced (typically several hours prior to the start time [12]), the parameters are used to determine route and ground delay assignments for the AFP-affected flights. We assume that airlines are implicitly incentivized to provide their most up-to-date cost parameters in order to maximize their likelihood of obtaining desirable flight plans in the AFP. This model does not employ means of providing additional incentives or equity in resource rationing.

This model is formulated as a route assignment problem with a system optimal solution objective. The outcome of this model will be the number of flights, X_r , to assign to each route, r , to minimize the total cost of AFP to operators. It is assumed that AFP flights are in competition for the available resources of lowest cost. As the AFP departure time increases for each subsequent flight i assigned to route r , and $g_{0,i,r} = 0$, the ground delay of flights on a route is monotonically increasing.

2) Model Specification

The N total flights originally scheduled to fly nominal Route 1 (Figure 3) in T are reassigned to one of R routes with new departure times. We assume that the flight operators submit different en route cost parameter values to traffic managers, such that $\alpha_1 \neq \alpha_2 \neq \dots \neq \alpha_N$. We assume that cost parameters are distributed over the flight population according to a probability distribution, and the N AFP flights are a representative population sample. Furthermore, if N flights are ordered by increasing α , we define $\alpha(n) = \alpha_n$ to be the en route cost parameter for the n^{th} flight. The value of α_n is determined as shown in the left graph of Figure 3.

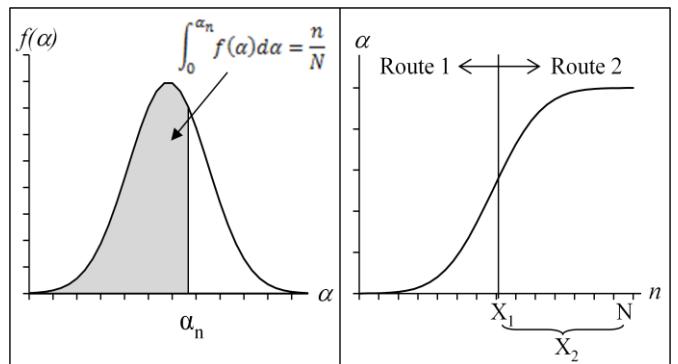


Figure 3. PDF of En Route Cost Parameter α across Flights

Given a set of routes, flights with the highest α values should be assigned to the routes with lowest en route times, and vice versa, if the unique minimum cost solution is to be obtained. The right graph of Figure 3 shows α plotted over n (shown as a continuous variable). For instance, if there are two routes such that $\rho_1 > \rho_2$, flights with lower α should be assigned to Route 1 such that those with higher α can take Route 2. If there are more than two route options, we order them according to decreasing en route times $\rho_1 > \rho_2 > \dots > \rho_R$, and aircraft can be ordered and assigned by increasing α .

We assume that α is uniformly distributed in $(\alpha_{\min}, \alpha_{\max}]$. Then α is a linearly increasing function of n such that:

$$\alpha = \alpha_{\min} + \left(\frac{\alpha_{\max} - \alpha_{\min}}{N} \right) \cdot n \quad (4)$$

The model is defined as follows:

Decision variable: $X_r \forall r$ (total flights assigned to route r)

Objective function (as per Equation (3), with $g_0 = 0$):

$$\min_{X_1, \dots, X_R} \hat{C} = \sum_{r=1}^R \sum_{i=1}^{X_r} \left(\rho_r \cdot \left[\alpha_{\min} + \theta \cdot \left(\sum_{j=1}^r X_{j-1} + i \right) \right] + g_r i \right) \quad (5)$$

where $\theta = \frac{\alpha_{\max} - \alpha_{\min}}{N}$.

Constraints: $\sum_{r=1}^R X_r = N; X_r \geq 0, \forall r$

The first part of the objective function represents the cost of additional en route time for flight i on route r , while the last term represents the ground delay for flight i on route r . The terms within the square brackets represent the α value for i on r . The choice of which flights to assign to which routes is based on the ordering described before Figure 3. The X_1 flights with the highest α values are assigned to route 1; the X_2 flights with the next highest α values are assigned to route 2; and so on. One can see that given the X_r values, this will yield the lowest cost assignment.

The first constraint ensures that all the flights caught in the AFP will be assigned to an available route and departure slot. The second constraint ensures that all route counts are non-negative. The objective function was checked for convexity. X_r is an integer variable, but this was relaxed to find a solution. Even if solutions are not integer, rounding (to preserve N) will still produce acceptable results because the headways on each route should be designed include some buffer space [5]. Also, if by rounding up X_r the route capacities were slightly exceeded occasionally, it would not be catastrophic.

If $X_n < 0$ or $X_n > N \ \forall n$, then interior solutions to the objective function of (6) do not exist, and solutions lie at the boundaries. In these cases, $X_n^* = 0$ and $X_n^* = N$ respectively.

Recall that the resulting resource allocation scheme is based on the estimated costs to operators. If $\varepsilon_{i,r}$ represents the unknown cost component for a flight, the “true” cost of the scheme is calculated by adding an error term to the total cost.

$$C = \sum_{r=1}^R \sum_{i=1}^{X_r} (\alpha \cdot \rho_r + g_r i + \varepsilon_{i,r}) = \hat{C} + \sum_{r=1}^R \sum_{i=1}^{X_r} \varepsilon_{i,r}, \quad \varepsilon_{i,r} \sim P \quad (6)$$

If $\varepsilon_{i,r}$ are iid Gumbel with parameters (a, b) , then according to the central limit theorem their sum ε is asymptotically distributed normal with mean $a - 0.5772b$ and standard deviation $\frac{\pi}{\sqrt{6}}bN$. We use $E[\varepsilon] = a - 0.5772b$ in the analytical solution. For simulated solutions we sample $\varepsilon_{i,r}$ N times to find C .

The Gumbel distribution has several important properties that make it analytically convenient to use in the specification of choice probabilities and expected cost [13], which we utilize for one of the stated route preference models that are discussed next. Also, the Gumbel distribution is reasonably similar to the normal distribution.

C. Stated Route Preference Model

1) Concept

The stated route preference models utilize the FCAR delay threshold concept [12]. FCAR was developed in order to give operators flexibility in identifying the best reroute options for their AFP-impacted flights.

In the FCAR process, operators of impacted flights are asked to submit route preference information to the traffic managers. For each route r , the operator of flight n submits a delay threshold value, $\Delta_{n,r}$, the cost at which flight n should be switched from route r to another. The quantity $\Delta_{n,r}$ contains the airlines’ complete cost information about route r , relative to the original flight plan, before ground delays are assigned. $\Delta_{n,r}$ is expressed in units of ground delay minutes such that airline costs are not explicitly revealed. Once the FAA receives the delay threshold values, they will rank flights route/departure time slot combinations based on some adopted resource rationing scheme [12]. For each sequential flight they choose a feasible departure time slot on each route, and based on the delay thresholds, determine the flight’s minimum cost route.

An example is shown in the figure below. Suppose a flight n had three route options, and the flight operator submitted a delay threshold value for each route r ($\Delta_{n,r}$). Once it is that flight’s turn for allocation, traffic managers check the slot availability on each route and determine the ground delay that flight n must take on each route: $GD_{n,1}$, $GD_{n,2}$, or $GD_{n,3}$. The route assigned to flight n is the lowest cost route, or $\min(\Delta_{n,1} + GD_{n,1}, \Delta_{n,2} + GD_{n,2}, \Delta_{n,3} + GD_{n,3})$. According to the figure this is route 3.

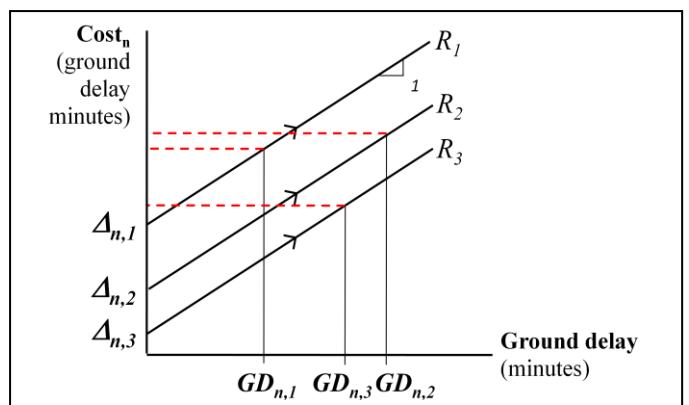


Figure 4. Delay Thresholds

We consider two stated route preference model scenarios. In the first, an AFP has been announced, and FAA traffic managers request flight operators to submit their delay threshold inputs by a deadline. Resources are allocated only after this deadline, when traffic managers have presumably received most or all flights' information. To represent this procedure we employ a model where the entire set of inputs is considered simultaneously in making allocations. We then consider a second system where flight operators are allocated their preferred resources on a first-submitted, first-assigned (FSFA) basis. It is envisioned that operators would be incentivized to submit their inputs as soon as they are able.

In the stated route preference model we assume that each airline would calculate the additional cost of a reassignment option using (2). However, airlines do not know what slots the FAA has available for their flight(s) on each route, and therefore have no information about the amount of ground delay that will be assigned to their flights. As a result, based on the flight cost model as specified in (1), airlines will submit delay thresholds ($\Delta_{n,r}$) that are calculated as follows. Traffic managers use these to compare the cost of route options combined with different ground delay slots.

$$\Delta_{n,r} = \alpha_n \rho_r + \varepsilon_{n,r}, \quad \varepsilon_{n,r} \sim P \quad (7)$$

Our specification assumes that a delay threshold is the airline's "true" and complete generalized cost for a flight n to fly route r before ground delay is assigned. Traffic managers will allocate resources to each flight through a particular allocation scheme using these delay thresholds. The delay thresholds ensure that under any combination of ground delay slots that could be assigned to their flight, the airlines have informed the FAA about which resources are of maximum utility to them.

2) Batch Model

In the batch model it is assumed that traffic managers receive delay thresholds from all airlines with AFP-impacted flights, before allocating resources, such that there still is no reward for submitting delay thresholds earlier than others. Airlines do not optimize or choose any resource options by offering delay thresholds; they simply offer the requested information about each of their choices to the FAA for use in the optimization. As a result the model remains a route assignment problem with a system optimal solution. The batch model is formulated identically to the parametric models except that the error term is included in the objective function, to represent the fact that airlines submit complete information about their preferences through their delay thresholds.

Again assume we have the situation of Figure 2 where N identical flights are to be reassigned to one of R routes with departure slots $d_{n,r}$. We want to know how many flights should be assigned to each route to minimize total user cost.

Decision variables: $X_r, \forall r$

Objective function:

$$\min_{X_1, \dots, X_R} C = \sum_{r=1}^R \sum_{i=1}^{X_r} (\Delta_{i,r} + g_r i), \quad \varepsilon_{i,r} \sim P \quad (8)$$

where $\Delta_{i,r} = [\alpha_{min} + \theta(\sum_{j=2}^r X_{j-1} + i)] \cdot \rho_r + \varepsilon_{i,r}$, and

$$\theta = \frac{\alpha_{max} - \alpha_{min}}{N}$$

Constraints: $\sum_{r=1}^R X_r = N; X_r \geq 0, \forall r$

Condition: Order routes such that $\rho_1 > \rho_2 > \dots > \rho_R$; order flights by increasing α .

Because this model contains random variables unique to each flight and route (i.e. the error term is contained in the objective function), Equation (8) cannot be solved analytically. However, we can treat each flight as an individual entity, where the decision variables are binary indicators of the route that each flight chooses. The model is formulated as binary integer quadratic program (BIQP) where the CPLEX solver is used to obtain a solution using the branch and bound algorithm. The results of this model tell us what route each individual flight is assigned to. Let's say that

$$x_{n,r} = \begin{cases} 1 & \text{if route } r \text{ is chosen for flight } n \\ 0 & \text{otherwise} \end{cases}$$

Decision variables: $x_{n,r} \quad \forall n, r$

Objective function:

$$\min_{x_{n,r}, \forall n, r} C = \sum_{n=1}^N \sum_{r=1}^R \left[\Delta_{n,r} + g_r \cdot \sum_{k=1}^n x_{k,r} \right] \cdot x_{n,r} \quad (9)$$

where $\Delta_{n,r}$ and θ are as defined previously.

Constraints: $x_{n,r} \in \{0,1\} \quad \forall n, r, \sum_{r=1}^R x_{n,r} = 1 \quad \forall n$

Constraint 1 restricts $x_{n,r}$ to be binary; constraint 2 ensures that each flight has been assigned to one route. The matrix for $\varepsilon_{n,r}$ was built from $N \times R$ random draws of the Gumbel distribution.

3) First-submitted, First-assigned (FSFA) Model

In the first-submitted, first assigned (FSFA) model, FAA traffic managers receive delay thresholds from operators in a sequence unknown beforehand. Each time an operator submits their delay thresholds for a flight, they are allocated the best possible resources available at the time, without considering future requests. This is identical to each flight choosing the minimum cost route and slot combination available. As a result, the FSFA process can be represented using the log-sums concept of the logit discrete choice model [13]. When the unknown portions of the utilities are assumed to be iid Gumbel with location parameter a and scale parameter b , the expected minimum cost and choice probabilities associated with a set of alternatives can be found. According to [14] and [15], the probability of agent n choosing an alternative r is:

$$P(V_{n,r}) = \frac{\exp\left(\frac{1}{b} \cdot V_{n,r}\right)}{\sum_{j=1}^R \exp\left(\frac{1}{b} \cdot V_{n,j}\right)} \quad (10)$$

where $V_{n,r}$ is the deterministic utility of r to agent n . In choice modeling we are typically concerned with the cost difference between two alternatives. If $E[W_n]$ is the expected cost of an alternative to n and $E[c_n^0]$ is that of another, then the difference between the two is:

$$E[c_n] = E[W_n] - E[c_n^0] \\ = \frac{1}{\gamma_n} \left(b \cdot \ln \left[\sum_{r=1}^R \exp \left(\frac{V_{n,r}}{b} + a \right) \right] - b \cdot \ln \left[\exp \left(\frac{V_n^0}{b} + a \right) \right] \right) \quad (11)$$

where W_n is the cost to operator n , γ_n is the (constant) marginal utility of income, and a and b are the Gumbel distributional parameters.

In the context of the AFP assignment, $E[c_n]$ represents the additional expected cost for flight n due to the AFP. We represent the deterministic utility using the cost function for a flight in the AFP such that

$$V_{n,r} = -(\alpha_n \cdot \rho_r + d_{n,r}), \quad V_n^0 = 0, \forall n, r \quad (12)$$

Recall that $d_{n,r}$ is the departure time (and ground delay, since scheduled departure times are $t \approx 0$ for the formulations introduced in this paper) for flight n assigned to r . Since the utility function $V_{n,r}$ is represented directly by the cost equation, $\gamma_n = 1$. We rewrite Equation (11):

$$E[c_n] = b \cdot \ln \left[\sum_{r=1}^R \exp \left(-\frac{\alpha_n \cdot \rho_r + d_{n,r}}{b} \right) \right] \quad (13)$$

The location parameter a cancels out of the equation due to its inclusion in the AFP cost and in the original cost. We now describe the recursive procedure by which the expected minimum cost is calculated for each flight.

- 1) Assign α_n value to each flight n . Randomly order the flights to simulate their unknown submission order.
 - 2) For flight $n = 1$, we calculate $V_{1,r}$, $P_1(r)$, and $E[c_1]$ using (12), (10), and (13) respectively, for all r .
 - 3) For $n = 2, 3, \dots, N$:
 - a. Determine the expected ground delay $E[d_{n,r}]$ on each route r for flight n . $E[d_{n,r}]$ is calculated based on the conditional probability that the previous flight ($n-1$) took r . Event “($n-1$) took route r ” is represented by B ; event “($n-1$) did not take route r ” is represented by $(1-B)$. $E[d_{n,r}]$ then becomes:
- $$\begin{aligned} E[d_{n,r}] &= E[d_{n,r}|B] \cdot P(B) + E[d_{n,r}|(1-B)] \cdot (1-P(B)) \quad (14) \\ &= (E[d_{n-1,r}] + g_r) \cdot P(B) + E[d_{n-1,r}] \cdot (1-P(B)) \end{aligned}$$
- $P(B)$ is the probability of agent $n-1$ taking route r , and was calculated in step 2 using (10).
- b. Find the expected utility of each alternative route for n , expressed as $E[V_{n,r}] = \alpha_n \rho_r + E[d_{n,r}]$.
 - c. Calculate the expected cost $E[c_n]$ using (13), using $E[V_{n,r}]$ calculated in (b).
 - d. Find the route choice probabilities $P(V_{n,r})$ as in (10), using $E[V_{n,r}]$.
 - e. Repeat (a) through (d) until $n = N$.
- 4) Find $\sum_{n=1}^N E[c_n]$.

We can perform the above calculations for different values of the Gumbel scale parameter b , where increasing b increases the variance of the Gumbel-distributed error term $\varepsilon_{n,r}$.

V. NUMERICAL EXAMPLE

When the FAA has perfect information ($\varepsilon_{n,r} = 0 \forall n, r$), the parametric (P1) and batch stated route preference (SP1) models are identical and hence yield identical resource allocations and total costs. As the traffic managers’ uncertainty about the airlines increases, the P1 cost result should remain the same, as resource allocations do not take the (changing value of) error into account. The SP1 model uses complete information to do a system-optimal resource allocation; as such, it will always yield the minimum total cost solution under any error variance. For this reason the SP1 solution is the baseline result. Under a zero error assumption, the FSFA stated preference model (SP2) solution will be equal or inferior to the other models because it does not offer a system-optimal solution. With greater uncertainty we might expect the total cost of the SP2 solutions to decrease like that of SP1.

To obtain insight into the performance of the three models under increasing uncertainty, which we model using increasing error variance, we present a numerical example. Suppose $N = 200$ flights must be reassigned routes and departure times as part of the AFP. The nominal route remains open but with reduced capacity. There are a total of 5 routes to which flights can be reassigned; the details are contained in Table 1. We consider the scenario where air cost ratios α are evenly distributed between (1,25] across the N flights.

As the interest is in relative rather than absolute performance, Figure 5 shows the cost differences of P1 and SP2 against the cost of SP1. SP1 requires simulation of the error term, and the results shown below are for 10 iterations.

There are three important conclusions to make from Figure 5. Firstly, as the FAA knows less and less about the airlines, the parametric (P1) model solutions degrade in comparison to those of SP1 and SP2. Secondly, the cost difference between the SP1 and SP2 results is consistent over increasing error variance. This is due to the fact that the error is known in both the SP1 and SP2 decision making processes. Finally, one can observe that the P1 solution is superior to the SP2 solution when the traffic managers know more about the operators (i.e. at small variance levels). However, after a certain error level (a standard deviation of about 10% of the zero error cost solution) the SP2 solution is more cost efficient. This result is intuitive; when traffic managers have plentiful and accurate information

TABLE I. SCENARIO FOR NUMERICAL EXAMPLE

Route	Capacity (aircraft per hour)	Departure Headway, g_r (min)*	En Route Time, h_r (min)	ρ_r (min)
1	30	2	125	25
2	12	5	120	20
3	7.5	8	110	10
4	6	10	107	7
5 (nominal)	4	15**	100	0

* This is the arrival (and departure) headway at Fix A.

** Headways after capacity is reduced due to AFP.

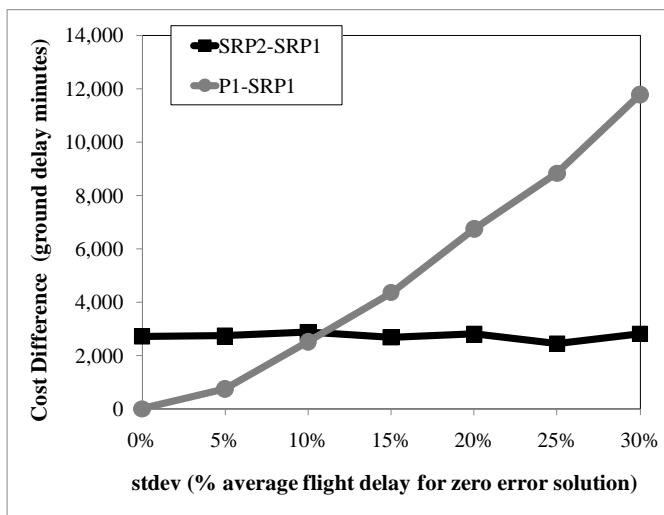


Figure 5. Total Cost, Numerical Example

about the airlines, the system optimal resource allocation will be superior to the FSFA allocation using complete information. However, when traffic managers have less information about the airlines, it becomes better to do a FSFA allocation with complete information rather than a system optimal allocation with incomplete information. Identifying these types of trade-off points is the core of this research.

Numerical checks demonstrated that the parameter values ($\alpha, \alpha_n, g_n, \rho_r$) have little effect on the solutions' relative positions to one another. Formal sensitivity tests will be performed as part of future work.

VI. DISCUSSION & FUTURE WORK

In this paper we propose a modeling framework through which we can investigate the many issues involved with incorporating user inputs in allocating constrained airspace capacity. We develop, evaluate and compare three user input and resource allocation schemes, under differing assumptions about how much traffic managers know about airline flight costs. The numerical example demonstrated the situations under which better information quality could be more desirable than timeliness, and vice versa. Building a model framework through which we can identify these types of tradeoff points is a key contribution of this research effort.

There are several important questions that are, and will continue to be, addressed. How much are flight operators willing to sacrifice input quality in order to submit their inputs faster? How does the timing of traffic managers' decisions affect the quality of their decisions to the operators? Also, airlines update their information constantly in the GDP and AFP databases. Given that their objectives and goals change so continually and rapidly, how will this affect decision-making when the goal is to maximize their utility? Addressing these questions is central to this research effort. As a result, it is important to continue discussions with practitioners, in order to better understand and represent airline behavior within the modeling framework of this research.

As part of on-going work, we are developing an additional stated route preference model, consisting of a hybrid between the system-optimal and FSFA resource allocation schemes. The advantage of this model is that it preserves the FSFA reward structure but potentially offers greater cost efficiency. We would like to develop a performance assessment procedure that combines user cost metrics with traditional operational performance metrics and emissions metrics. We would also like to improve the user cost specification by including missed connections, to account for downstream effects of flight delay.

This research investigates the interaction and information exchange between flight operators and the FAA. The ultimate goal is to provide insight into the potential mechanisms of collaborative resource allocation within the context of the AFP, in order to guide future AFP policy decisions.

ACKNOWLEDGMENT

The authors would like to thank the FAA for sponsoring this work. We would also like to thank Bob Hoffman, Dennis Gallus, and Nathaniel Gaertner at Metron for their great assistance, insights, and suggestions.

REFERENCES

- [1] Taber, N. J. (2005). *Operational Concept for Integrated Collaborative Rerouting (ICR)*. McLean, VA: The MITRE Corporation, MTR05W000053.
- [2] Bertsimas, D., & Stock Patterson, S. (1998). The Air Traffic Flow Management Problem with Enroute Capacities. *Operations Research*, 46 (3), 406-422.
- [3] Goodhart, J. (2000). *Increasing Airline Operational Control in a Constrained Air Traffic System*. University of California, Berkeley: PhD Dissertation.
- [4] Andreatta, G., & Romanin-Jacur, G. (1987). Aircraft Flow Management Under Congestion. 21.
- [5] Richetta, O., & Odoni, A. R. (1993). Solving Optimally the Static Ground-Holding Policy Problem in Air Traffic Control. *Transportation Science*, 228-238.
- [6] Ball, M. O., Hoffman, R., Odoni, A., & Rifkin, R. (2003). A Stochastic Integer Program with Dual Network Structure and its Application to the Ground-Holding Problem. 51 (1).
- [7] Jakobovits, R., Krozel, J., & Penny, S. (2005). Ground Delay Programs to Address Weather within En Route Flow Constrained Areas. San Francisco, CA: AIAA Guidance, Navigation and Control Conference and Exhibit.
- [8] Vossen, T., Ball, M., Hoffman, R., & Wambsganss, M. (2003). A General Approach to Equity in Traffic Flow Management and its Application to Mitigating Exemption Bias in Ground Delay Programs.
- [9] Hoffman, R., Burke, J., Lewis, T., Futer, A., & Ball, M. (2005). Resource Allocation Principles for Airspace Flow Control. San Francisco: AIAA Guidance, Navigation, and Control Conference and Exhibit.
- [10] Mayet, J., & Hansen, M. (2000). Congestion Pricing with Continuously Distributed Values of Time. *Journal of Transport Economics and Policy*, 34, Part 3, 359-370.
- [11] Ball, M., Futer, A., Hoffman, R., & Sherry, J. (2002). Rationing Schemes for En-route Air Traffic Management (CDM Memorandum).
- [12] Hoffman, R., Lewis, T., & Jakobovits, R. (2004). Flow Constrained Area Rerouting Decision Support Tool, Phase I SBIR: Final Report.
- [13] Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press.
- [14] Domencich, T. A., & McFadden, D. (1975). *Urban Travel Demand: A Behavioral Analysis*. Amsterdam: North-Holland.
- [15] Ben-Akiva, M., & Lerman, S. R. (1994). *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press.

An optimisation framework for aircraft operators dealing with capacity-demand imbalances in SESAR

Luis Delgado Xavier Prats
 Technical University of Catalonia (UPC)
 Castelldefels (Barcelona), Spain
 Emails: [luis.delgado, xavier.prats]@upc.edu

Abstract—This paper presents a framework for the negotiation phase that is foreseen in the new operational concept proposed in the Single European Sky Research (SESAR) program. In particular, this paper describes a possible strategy for the airspace users in order to deal with the Collaborative Decision Making (CDM) process that is expected in this future scenario. In the SESAR scenario, airspace users will become owners of their trajectories and they will be responsible to solve possible mismatches between capacity and demand in a particular airspace sector. The aim of this strategy is to improve the efficiency in the CDM process by computing the different operational costs associated to different solutions that may solve a particular demand-capacity imbalance in the airspace. This will allow them to optimise their operating costs while reducing fuel consumption and therefore being more environmentally friendly. Some suggestions have already been done for the CDM mechanism, for instance the use of auctions. However, the different options that aircraft operators might use have not yet been sufficiently investigated. In this paper, the authors propose an optimisation framework for aircraft operators aimed at computing 4D trajectories with time constraints to deal, in this way, with possible airspace regulations. Once a nominal flight plan and a potential regulation is known, it is suggested to compute several possible alternative flight plans (including rerouting, but also altitude and speed profiles) that may solve the capacity-demand problem. If more than one regulation is applied to the flight, a tree of options is subsequently computed. The cost of each optimised the option is also calculated in order to allow the airspace users to initiate the negotiation process with other airlines. Finally, a preliminary example is given at the end of this paper in order to better illustrate the proposed methodology.

I. INTRODUCTION

As it is well known, the number of IFR flights is growing all around the world. The forecast of flight movements in the Eurocontrol Statistical Reference Area (ESRA) for 2030 is between 1.7 and 2.9 times the traffic of 2007 [1] and, according to [2], by 2030 the 11% of actual demand will not be accommodated, in the most-likely growth scenario. For example, during the period from 2003 to 2008, the European traffic has increased by 19.9% (average of 27818 flights per day in 2008), the total delay has increased by 60.7% (65138 minutes per day) and the total delay per flight has increased by 34% (2.3 minutes on average for all flights) [3]. This trend shows that capacity of the system is starting to get overpassed. To deal with capacity-demand imbalances, ground delay programs have been implemented. The ground holding problem has been thoroughly addressed in all its forms: as a deterministic process [4], probabilistic [5], [6], for a single airport [4] or for a multi-airport scenario [7]. For instance

in [7] the multi-airport ground holding problem is solved using integer programming techniques. The objective of these algorithms is to determine the ground delay that has to be assigned to flights in order to deal with capacity constraints. However, all these models need accurate information about the flights, and in particular, the costs associated to delay.

On the other hand, techniques as the one described in [8] allow to analyse the propagation of delay on a network of more than one airport. These techniques that are focused on the airport have been extended in order to deal with all the network constraint, including airspace capacity restrictions. In this manner the whole Air Traffic Flow Management (ATFM) problem, with ground delay, speed control during cruise and rerouting, has been solved (see for instance, [9], [10] or [11]). For a wider and excellent literature review of modelling and optimisation in traffic flow management, the reader is referred to [12].

Even if all previous approaches are able to compute the best route, the optimal amount of ground delays and even the optimal cruise speed for the different flights, these computations are done in a centralised system aimed at optimising the whole network. The main assumption for this system is that it is supposed to be fed with accurate data coming from the aircraft operators. Yet, some of the data are considered *critical* for the airlines, specially when dealing with cost figures, and they would be reluctant to release them. In other words, keeping the problem centralised, the above techniques are appropriated to solve it, but some effort has to be done to include airlines preferences while maintaining the privacy of some of their data. Nowadays, priority has been given to user-driven policies and therefore, as traffic is expected to continue growing, new concepts of operation are starting to be developed: SESAR project (in Europe) and NextGen (in the USA).

If the focus is given to Europe, two big changes arise from the SESAR guidelines: 4D trajectories should become a reality and the airspace users (i.e. the aircraft operators) will be the owners of their trajectories. The ownership of the trajectories leads to a situation where if a capacity-demand imbalance exists, a negotiation process among airlines should be done to solve the potential conflicts. The network managers are not longer in charge of solving the imbalance in a centralised manner but of coordinating the negotiation between the airspace users. The airspace users will be involved in the process of balancing demand and capacity and a Collaborative Decision

Making process (CDM) will become mandatory at strategic level [13]. During summer 2008, 14.1% of the traffic in Europe was delayed with an average delay of almost 20 minutes [14]. Furthermore, in 2008 the price of fuel reached values over \$100 per barrel and therefore, most airlines reported fuel costs to be between the 30 and 40 percent of their total expenses. Therefore, on one hand, the aircraft operator will be forced to deal with capacity-demand imbalance, while on the other hand, bearing in mind that the objective of the aircraft operator is to improve its benefits, optimise its 4D trajectories according to the cost of time and fuel burned. An optimisation is essential if they want to reduce their operational costs and therefore, be more competitive in front of other operators.

In the future SESAR scenario, it will be critical for airlines to know the associated cost of solving capacity-demand imbalances in the air transportation network. Therefore, if a negotiation process is established with concurrent airlines, those ones with more options, and with better information of the associated costs for each option, will be better placed [15]. In this context, the negotiation process has already been analysed in [16], where a market based mechanism is suggested to be used. However, the different options that the aircraft operators would have when facing this negotiation process have not been yet assessed and this is the main motivation of the proposed research by the authors.

Thus, this paper suggests an optimisation framework for aircraft operators that have to negotiate with other airlines in order to solve a capacity-demand imbalance problem in the airspace. In this negotiation process, different slots might be traded. In this case, it would be essential for the airline to compute the optimal vertical profiles and speeds to be used for each of the possible options, that will result in a different final costs. When a regulation is set, the affected airspace users will initiate the negotiation process acting in different ways, which represent different options, according to their own interests and to the associated costs of each solution. Therefore, the proposed methodology is intended to assess the different options that a particular aircraft operator would have and to compute the associate cost for each of them in order to better perform in the negotiation.

This paper is organised as follows: in Section II the current framework of operations used in Europe is presented, regarding both the the network manager and the airlines. Section III presents the operational framework in the SESAR scenario while taking into account the proposal of the authors for the aircraft operators. Section IV is devoted to show a preliminary example of the proposed methodology, considering the computations that a given airspace user would perform for a hypothetical regulation. Finally, in Section V the main concepts are summarised and further work on this research is explained.

II. CURRENT OPERATIONAL FRAMEWORK

Nowadays, in the operational concept, as implemented in Europe, the Air Navigation Service Providers (ANSP) submits the capacity of their airspace sectors to the Control Flow

Management Unit (CFMU). The CFMU acts as a network manager and has the responsibility of maintaining the demand below the capacity for each sector. In order to attain this objective, the airspace users must submit their intended flight plans to the CFMU well in advance. As it can be seen in Figure 1, the CFMU regulates the demand by imposing on-ground delays to some of the flights.

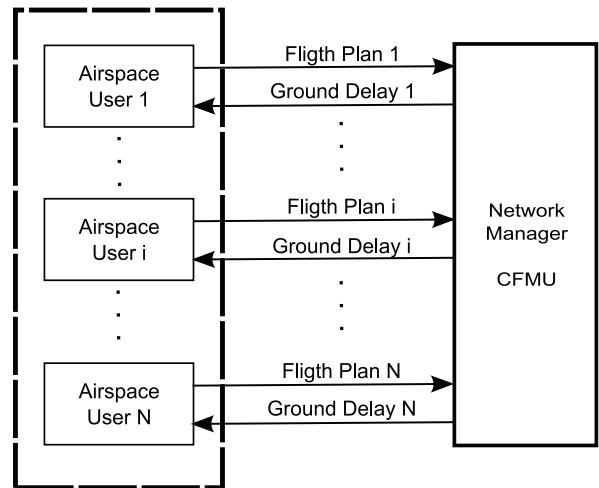


Fig. 1. Current concept of operations in Europe

On the other hand, airline operators optimise their flight plan with respect the cost of time and fuel. During this optimisation process different operational parameters are taken into consideration, such as crew and maintenance costs, number of transfer passengers, the type of the aircraft, weather conditions, available airspace routes, etc. However, airspace capacity information is hardly never taken into account. Next, current airspace network management and airline operation strategies are briefly described.

A. Network Manager

In Europe, the CFMU simulates the flight plans in order to identify those sectors where the capacity might be exceeded by the foreseen demand. In this case, the Computer Assisted Slot Allocation (CASA) algorithm is used to mitigate this mismatching by imposing on-ground delays to some flights. CASA implements a First Planned First Served (FPFS) algorithm to assign slots to flights while preserving fairness. Briefly, this slot allocation algorithm can be explained by the following simple example.

Let us set a regulated area with one available slot every five minutes (10:00, 10:05, 10:10...), and six planes that want to cross this regulated airspace with their Estimated Time of Over-fly (ETO) as shown in Table I. As it can be seen in Figure 2, the first plane (F1) will take slot number one while F2 will take slot number two. Without any regulation, the ETO of the third aircraft (F3) is 10:07, corresponding as well to slot number two (between 10:05 and 10:10). However, this slot has been already assigned to F2 that will keep it as its ETO is lower than the ETO of F3. Then, the third slot will be

TABLE I
EXAMPLE FLIGHTS

Flight	ETO
F1	10:00
F2	10:06
F3	10:07
F4	10:10
F5	10:12
F6	10:18

assigned to F3 and this flight will be delayed on ground by three minutes. In the event of having more than one regulation, the delay coming from the most penalising regulation will be imposed to the aircraft. Then, the over-flight time of the remaining regulations will be fixed to this most restrictive value [17].

The final result that is obtained with this assignation is shown in Figure 2. As it can be seen, flight F3 has been delayed for three minutes, and will arrive at the regulated area at the slot R1S3, flight F4 will be delayed for five minutes and will use slot R1S4. Finally, F5 would have arrived at the regulated area to take slot R1S3, but being the CASA algorithm FPFS, it must be delayed DGF5 minutes in order to arrive at the regulated area with slot R1S5. In Figure 2, the slot that F5 would have taken is presented along with the finally assigned one and the ground delay (GDF5) that consequently has been imposed to this flight. It is worth mentioning that besides the departure time, the flight plans of the delayed flights are not changed. This means that once the delay has been absorbed on ground, the flight will be operated at its initially planned cruise speed.

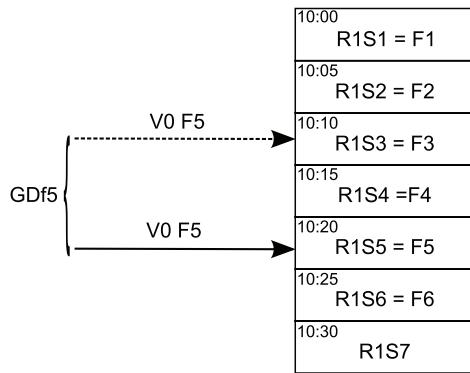


Fig. 2. Example of a regulation area with 5 slots every five minutes

The main advantages of this solution are that it is simple to find a robust solution, the algorithm can easily deal with real-time modifications and cancellations of flight plans and, being a FPFS algorithm, a minimisation of the total delay is achieved [16]. However, it does not take into account the cost for the operators and the repercussion of the imposed delay on that cost. In other words, the economical impact of the regulation is not minimised while the same amount of delay can indeed be more expensive for a given operator than for

another [16], [18].

Some effort has been done in order to try to improve this CASA algorithm using new techniques as constraint programming (see for instance [19]) or extending the ground delay to deal with conflict and not only with capacity-demand imbalances [20]. Moreover, other criterions rather than the FPFS algorithm have been analysed, like for instance the ratio-by-distance one [21]. Nevertheless, this modifications of CASA algorithm present some issues that stop their practical implementation. Even if the computation time has been significantly reduced, they still have difficulties to deal with real time modifications and cancellations of flight plans. Moreover, some of them have problems with equity and fairness.

B. Airspace Users

The main objective of aircraft operators is to minimise their operating costs. Therefore they will try to compute and fly an efficient flight plan. In Figure 3 it is presented the optimisation process that the airline should do for each of its flights. Before this optimisation, the airline has computed the route planning and the fleet and crew assignment. The reader is referred to [22] and [23] for more details on these processes, which are out of the scope of this paper.

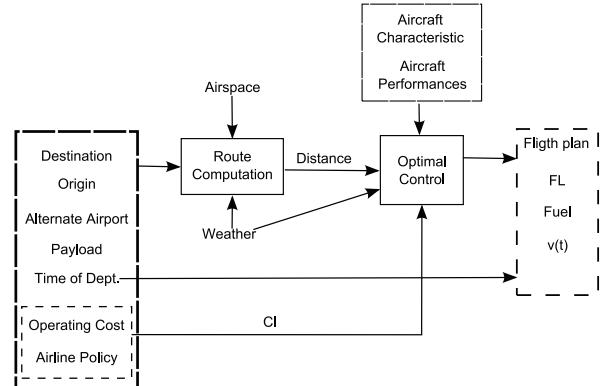


Fig. 3. Flight optimisation applied nowadays

In the flight plan optimisation, the input values are the route that the airline will fly (origin, destination and alternative airports), the intended payload and the time of departure. With the information of the airports and using the airspace configuration and the weather data, the route will be computed [24]. After this process, the distance to be flown will be obtained. A main aspect to take into account in this process is the airline policy regarding its operating costs. This will result on a given CI (Cost Index) for the intended flight. The Cost Index will be part of the optimisation function which will weight the cost of time against the cost of the fuel. Therefore, the optimisation function is to be $J = \text{Fuel} + CI \cdot \text{Time}$. As expected, changes on CI will impact on the profile of the flight, on the speeds and, as a result, on the fuel consumed and on the final take off weight [25]. It has been demonstrated how variations on CI might have an small impact on time but a great repercussion on fuel consumed [26].

Summing up, by using the aircraft characteristics and aerodynamic data, the payload, the distance, the weather and the CI, the optimiser computes the operational flight plan that is composed of speed and vertical profiles, as well as the fuel needed for that flight [27], [28].

During the flight, the CI is introduced in the Flight Management System (FMS) by the pilot. The management of the flight will be done by direct changes on the CI. This is the reason why it is not surprising that extensive research has been conducted to help airlines on the optimisation of their CIs. If a flight is delayed, but time is critical, which means that the cost of time is high, some time might be recovered during the flight. Nevertheless, as it has been analysed in [18], there is a compromise between the time recovered and the fuel burned. Therefore, to optimise the new value of CI becomes crucial [18], [29].

III. PROPOSED FRAMEWORK FOR SESAR

As mentioned before, the main change that SESAR introduces is that the airspace users become owners of their trajectories [13]. It means that in this new operational scenario, the network manager should not modify the intended flight plans of the aircraft, unless it is strictly necessary. In SESAR, as in NextGen too, the trajectories will be based on the 4D concept. A 4D trajectory is a precise description of the flight path of an aircraft as a 4 dimensional continuum, from its current position to the point at which it touches down at its destination. Thus, every point on a 4D Trajectory is precisely associated with a time [30]. Obviously, this will help on the predictability of the flights and some gain in efficiency is also expected. The airspace users will create their trajectories that in turn, will be shared using the network manager. With this information, along with the airspace related data, the airlines will have to negotiate among them to solve possible capacity-demand imbalances. In this case, the network manager will only act as a supervisor of the negotiation process that airspace users will do in case the demand excess the capacity (see Figure 4).

A. Network Manager

The task assigned to the network manager in the new operational context is the coordination of the different airspace users. As previously mentioned, in [16] a market mechanism aimed at assigning the air traffic flow management slots is proposed. In this case, after an initial First Planned First Served (FPFS) assignation (done by the network manager), an auction process is subsequently initiated. The airlines are owners of their initially assigned slots by the FPFS algorithm, but during the auction process they might keep or sell them according to their own interests.

In order to achieve an optimum from an economical point of view, the airspace users must have a good knowledge of the cost associated with a particular slot. This would help them to eventually choose a particular slot and sell their initial one, with regards to the other slots. In the work done by [16] and [18] a fixed cost is chosen for each minute of delay. In

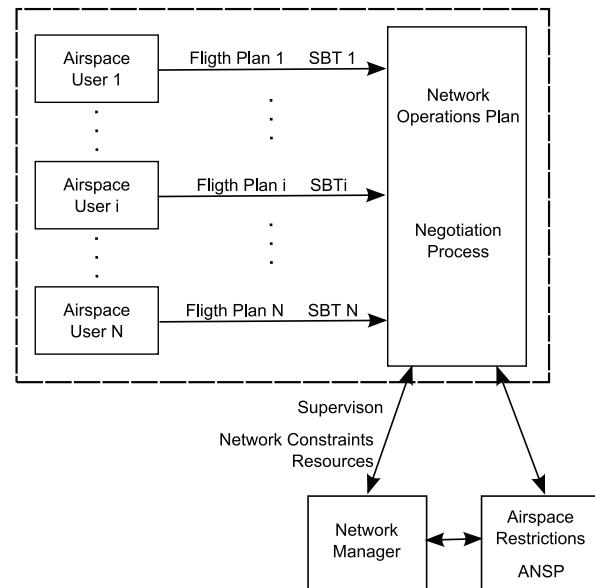


Fig. 4. SESAR concept of operations

these works, if the aircraft operator chooses a slot later than the initial one an extra on-ground delay must be performed (as shown in Figure 2) and no other options are left to the airlines. Moreover, in [16] the delay that the airline suffers at the take-off is supposed to be the same delay that the flight will experiment at the arrival airport. This means that the airline is not allowed to change the original flight plan that was proposed before the regulation was known. In addition, the possibility of speeding up the flight before the regulation is also not considered and therefore only the slots with a higher time than the one the aircraft would have without regulation are taken into account. However, as it will be shown in next section, the authors propose that airlines might be more active during the negotiation process. The aircraft operator should change the initial flight plan (i.e. vertical and speed profiles, or even re-routing) in function of the chosen slot.

B. Airspace Users

In a complete 4D environment, where airspace users can fully optimise their trajectories, many options arise to deal with capacity-demand imbalance problems. First, a re-routing may be possible in order to avoid the regulated area.

In the case where the original route is kept, the aircraft might take off later (as it is done nowadays with the on-ground delay methodology [17]) but it would be also possible to fly slower. In this way the aircraft would be in the air earlier and if for some reason the regulation is cancelled it would be easier for the operator to recover the initial delay. Moreover, by flying slower, the cost of arriving to a later slot is also optimised [29]. Finally, the aircraft could increase the cruise speed in order to arrive to a previous slot. In fact, the optimisation algorithm used by the airspace user should compute an optimal solutions for each possible slot taking into account a combination of all above strategies.

Once the regulation has been passed, some time might be recovered if the aircraft speeds up. Due to the fact that recovering time would have an impact on fuel consumption, in [18] an analysis has been done showing the amount of optimal time that should be recovered. As it could be expected, optimised solutions often do not recover all the possible delay due to the involved fuel consumption. On the other hand, even with high cost indexes, the time that is possible to be recovered is quite limited for short-haul routes. Thus, this technique may become more interesting for longer flights [18], [29].

The optimisation process that airspace users have to do will be enhanced to include time constraints, as shown in Figure 5. The authors propose the computation of the whole trajectory using an optimal control approach while meeting all possible constraints. Thus, the input of the optimiser will be, as in Figure 3, the distance computed, the weather conditions, the aircraft characteristics, but also the way-point time windows constraints for each slot and regulation.

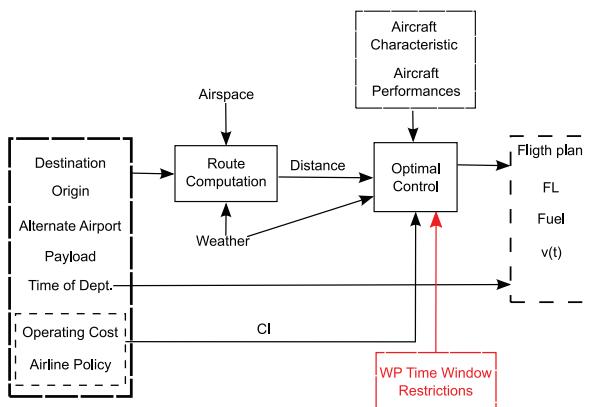


Fig. 5. Proposed flight optimisation

Therefore, each airline will compute for a given regulation a set of achievable slots. These sets will be bigger than other proposed approaches, such as [16], where all the delay is supposed to be absorbed on ground. The first valid slot will be determined by the aircraft taking off as soon as possible and flying to the regulated area at the maximum operational speed (or VMO). On the other hand, the last slot will be reached when flying at an optimal speed before the regulation to arrive at the slot ($V_{optAR_jS_i}$) and eventually doing some on-ground delay of GD_i . The last useful slot will be determined when the cost of the delay produced at the arrival airport due to the fact of using that slot becomes bigger than the economical profit that can be attained by using it.

After the regulation it will exist an optimal speed ($V_{optAR_jS_i}$) that will allow to eventually recover some time in order to minimise the cost of the delay at the destination airport. This optimal speed will take also into account the increase in fuel consumption due to the fact that the aircraft is flying faster than the initial intended speed [18]. The authors also suggest that the variable that should be taken into account in this optimisation process is the total delay at the destination

airport instead of the on-ground delay before take-off as it is usually done nowadays. In fact, the real cost for a minute of delay is due to the fact that the flight arrives late at the destination airport rather than because it has departed delayed.

In Figure 6, it can be seen that for each available slot, the airspace user will have a certain amount of ground delay (GD_i), an optimum speed to arrive to that particular slot ($V_{optAR1S_j}$) and another optimum speed after the slot to eventually recover or loose some time if necessary ($V_{optAR1S_j}$). These speeds should be computed with the optimisation mechanism proposed in Figure 5 by changing the time window associated to the way-point that define the entry of the regulated airspace. In Figure 6 it is shown that if the aircraft flies as initially planned, it will over-fly the regulated area at the slot achieved at V_0 . However, the aircraft operator has a set of alternative options, by using other slots with different associated costs on fuel and total delay. For each path (i.e. each different slot), the whole trajectory should be optimised by the aircraft operator. The optimal cost for each path will be computed in order to start the slot auction process described above. This optimisation might be done with an optimal control problem modelled with phases as the one described in [31], extended with time windows constraints.

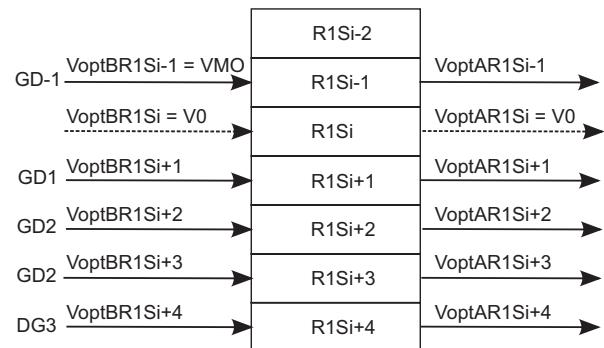


Fig. 6. One regulation with changes on flight plan

It is not surprising that the aircraft might fly through more than one regulated area. Actually in Europe 21% of the flights had two regulations in the AIRAC 311: 21st July 2008 to 27th August 2008 [16]. In this case, as can be seen in Figure 7, from one slot of the first regulation a set of slots on the second regulation can be reached flying from VMO to Vmin. After the second regulation an optimum speed ($V_{optAR2S_i}$) can be used to recover the optimal amount of time. Then, the optimiser has to be extended to include the possibility of having more than one restriction. This should not be difficult due to the fact that a narrow set of slots at the second regulation might be reached from one slot of the first one (see Figure 7).

Therefore, for each slot of the first regulation the airspace user have a set of slots of the second regulation that can be reached. With this definition a tree can be created (see Figure 8), and for each path different speeds will be used to minimise the operational cost (fuel and time). It is expected

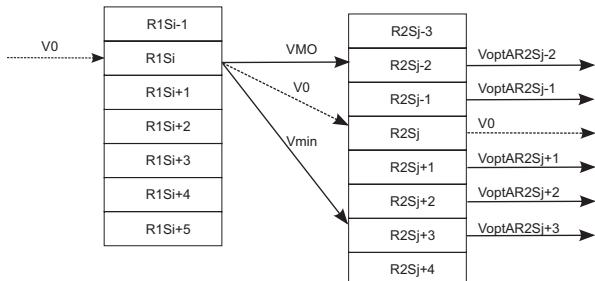


Fig. 7. Slots reachable from one slot of the first regulation

that this tree will not be too large, and therefore its creation should be computationally feasible. In this context, it has been presented in [18], [25] and [29] how time that can be saved or lost by changing cruise speed is quite reduced. Moreover, as it is distributed, each airline has to compute its own trees, reducing the computational cost that would involve solving this problem for all the traffic in a centralised system.

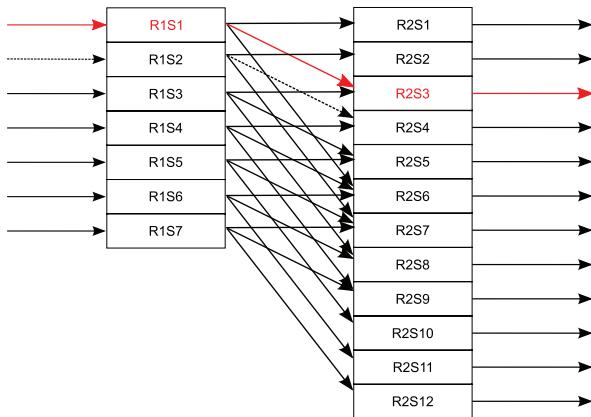


Fig. 8. Tree of reachable slots with two regulations in place

With this computation, the airspace user is able to determine which is the direct cost that it will have if a set of slots is chosen. If it is not possible to change the assigned slots, like in the current operational concept, the optimum speeds and vertical profiles to minimise the cost will be determined. If a negotiation process is possible with the network manager, the airspace user will be in a better position to choose between the options. Finally, if a market mechanism is established as the one described in [16] the airspace user that implements this solution will know the cost of each of the paths. Each path will be a set of slots, for example R1S1 and R2S3 which are shown in red in Figure 8. With the optimisation process, for each path the vertical profile and the optimum speeds will be computed. Therefore the airline that performs this optimisation has more information to decide at which price is worth to sell the original assigned slots and to buy a different path.

One advantage of this optimisation is that the objective functions for the airline can be easily modelled while the negotiation process supervised by the network manager ensures that the capacity is not exceeded. Moreover, the suggested

model allows to include different types of airlines, with different objectives and even airlines that do not optimise their trajectories with time constraints. The difference will be that those who did it will have more information and more optimal trajectories, being in a better situation to perform the negotiation [15].

The mechanism described in [16] might be easily extended to include re-routing. In this case, the airspace user will monitor the cost of different paths through different sectors while performing the negotiation.

IV. PRELIMINARY EXAMPLE

In this section, an illustrative example of the concept proposed above will be shown. The following preliminary results are based on a hypothetical situation where an Airbus A320 is scheduled to fly a route of 2000 NM with a payload of 15 tons. Let us suppose that the aircraft operator chooses a cost index (CI) of 40. For this aircraft and payload, this CI represents a cruise speed of M 0.789 with a total flight time of 250 minutes (the climb and descent phases are neglected in this preliminary example) [32], [33]. On the other hand, let us have a regulation located at 800 NM ahead from the departing airport and where airspace slots are available at six minutes intervals. For the sake of simplicity, the time references are set to zero at the original intended take-off time.

Figure 9 shows the initial intended flight plan, where the CI is set to 40. In this case, the aircraft will enter the regulated area after flying 107 minutes and therefore, it will use the third available slot (R1S3) that spans from minute 106 to minute 112. Let us assume that another flight with a lower ETO has already been assigned to this slot R1S3. This means, that our aircraft will be delayed for five minutes on-ground in order to enter the regulated area by using the slot R1S4. If the flight plan is not changed, as it is done nowadays, the aircraft will always fly at CI 40 and therefore will arrive to the destination airport with a delay of five minutes.

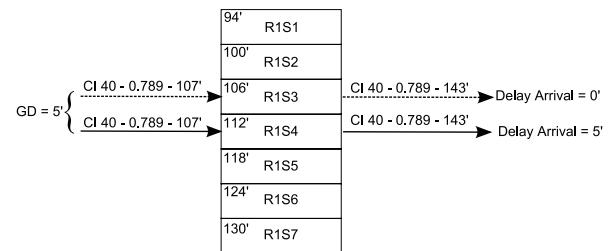


Fig. 9. Example of one regulated area without changes on the original flight plan

With the mechanism proposed in this paper, the aircraft operator can compute the cost of all available slots. For each slot a flight plan optimisation is performed in order to minimise their own policy of time and fuel consumption. Figure 10 shows the different available slots for this particular example. Even if the aircraft takes off at the original intended take-off time it is not possible to reach the regulation before 105 minutes, in order to arrive at slot R1S1, due to the

limitation on the maximum cruise speed. It turns that the first available slot for this example is the second one (R1S2), spanning from minute 100 to 106. To achieve this slot, no ground delay will be done and a CI of 150 will be used. For the studied aircraft this corresponds to a cruise speed of M 0.80 from the take-off to the regulated area. After the regulation it is possible to fly slower to save some fuel since the aircraft is two minutes ahead of the original schedule. In this case, the CI is changed to 25 and the flight will continue at M 0.78 during 145 minutes to the destination airport, where the plane will arrive on time.

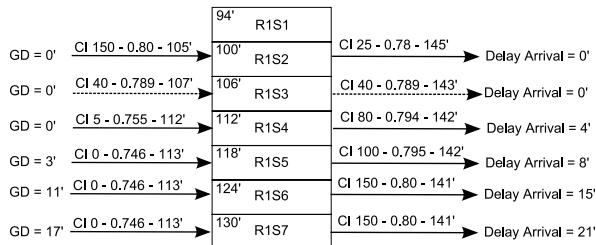


Fig. 10. Example of one regulated area allowing changes on the original flight plan

Obviously, for the third slot (R1S3) the flight is performed at the intended CI of 40 and any delay is experienced. If the slot R1S4 were to be used, it is worth mentioning that on the current operational scenario the aircraft would be delayed five minutes on ground (see Figure 9). However, with the proposed mechanism slot R1S4 can be reached with no delay on ground flying at a lower airspeed before reaching the regulation. In this case a CI of 5 would be used, corresponding to a cruise speed of M 0.755. Using this cost index, the plane will arrive to R1S4 consuming less fuel than initially planned, but with five minutes of delay. Moreover, once the regulation is passed, a speed up might be done by increasing the CI to 80. This will represent arriving with four minutes of delay instead of the initial five minutes expected with the current operational concept of operations. Finally, for the last three slots (R1S5, R1S6 and R1S7), the best that can be done is to fly at CI=0 to minimise the fuel consumption during the segment before the regulation while adding the needed ground delay in order to arrive to the regulated area at the appropriate slot. As it was done with slot R1S4, once the regulation is passed some time may be recovered speeding up the flight. In this case, the authors refer to the work presented in [18] where it is shown in which conditions it is worth to increase the airspeed by trading off fuel consumption and time recovered.

After this optimisation process, the aircraft operator knows exactly the cost associated to each slot, how much delay the flight would experience at the destination airport, how much fuel would be used and therefore the best sequence of CI depending of the flight segment. In this way, if a marked based mechanism is used, as described in [16], the airline will be on a better position to decide if it is worth to sell their initially assigned slot (in this example slot R1S4) and to buy another one.

V. CONCLUSION AND FURTHER WORK

This paper explains a framework for the optimisation of aircraft trajectories in the SESAR operational scenario, where airlines are expected to be more active in the resolution of demand-capacity imbalances in the airspace or airports. One possible concept of operations in the future Air Traffic Management (ATM) scenario is that if a congested airspace is declared, airspace users will have to agree with the final adopted solution. As airlines will have to negotiate the necessary delays or reroutings, a game from an economical point of view is set and the agent with most information is most likely to have advantage with respect to the others.

In this paper, the authors, suggest the idea of solving this kind of Air Traffic Flow Management (ATFM) problems in a distributed way, instead of using a centralised approach as it is often proposed in the literature. Therefore, airlines will accurately compute the cost of the different options that arise when the capacity-demand imbalance problem is solved (such as ground delays, speed reductions or reroutings). This optimisation will compute the optimal speed and altitude profiles for each possible alternative leading, in consequence, to different fuel consumptions and different delays at the destination airport. With the proposed methodology, this information will be kept by the operators without the necessity to publish sensible data to the network manager, as it is necessary with centralised based solutions. Thus, the main advantages of this method are that airliners can keep the secrecy on their data; the final solution is globally more efficient than with centralised methods, because airline data is expected to be more accurate; and being a distributed optimisation performed for each airline separately, no computational issues are expected when solving big real problems.

As part of on-going work, we are analysing the benefits of this solution with more than one restriction at the same time. In addition, results are being obtained for some practical cases using realistic data. Also, some simulations with the market mechanism are also foreseen including airlines with and without the optimiser in order to analyse the benefit for an airline of having this data available. Finally, as airlines work with the Cost Index (CI) parameter, a complete translation from this optimisation process to the CI values might be also interesting.

REFERENCES

- [1] Eurocontrol, *Long-Term Forecast Flight Movements 2008-2030*. Eurocontrol STATFOR, Statistics and Forecast Service, 2008.
- [2] ——, *Challenges of Growth 2008 - Summary Report*. Eurocontrol STATFOR, Statistics and Forecast Service, November 2008.
- [3] ——, *Network Operations Report 2008*. Eurocontrol, March 2009.
- [4] A. R. Odoni, "The flow management problem in air traffic control," in *In Odoni, A. R., Bianco, L., and Szego, G., editors, Flow Control of Congested Networks, pages*. Springer-Verlag, 1987, pp. 269–288.
- [5] M. Ball, R. Hoffman, A. R. Odoni, and R. Rifkin, "A stochastic integer program with dual network structure and its application to the ground-holding problem," *Operations Research*, vol. 51, pp. 167–171, 2003.
- [6] O. Richetta and A. R. Odoni, "Solving optimally the static ground holding problem in air traffic control," *Transportation Science*, vol. 24, pp. 228–238, 1993.

- [7] P. B. Vranas, D. J. Bertsimas, and A. R. Odoni, "The multi-airport ground-holding problem in air traffic control," *Operations Research*, vol. 42, no. 2, pp. 249–261, March-April 1994.
- [8] M. D. Peterson, D. J. Bertsimas, and A. R. Odoni, "Decompositon algorithm for analyzing transient phenomena in multiclass queueing networks in air transportation," *Operations Research*, vol. 43, no. 6, pp. 995–1011, November-December 1995.
- [9] D. J. Bertsimas and S. S. Patterson, "The air traffic flow management problem with enroute capacities," *Operations Research*, vol. 46, no. 3, pp. 406–422, May-June 1998.
- [10] ———, "The traffic flow management rerouting problem in air traffic control: a dynamic network flow approach," *Transportation Science*, vol. 34, no. 3, pp. 239–255, August 2000.
- [11] D. J. Bertsimas, G. Lulli, and A. Odoni, "The air traffic flow management problem: an integer optimization approach," in *Proceedings of the 13th Conference on Integer Programming and Combinatorial Optimization*, 2008, pp. 34–46.
- [12] B. Sridhar, S. R. Grabbe, and A. Mukherjee, "Modeling and optimization in traffic flow management," *Proceedings of the IEEE*, vol. 96, no. 12, pp. 2060–2080, December 2008.
- [13] S. Consortium, *The ATM Target Concept*. SESAR Consortium, September 2007.
- [14] Eurocontrol, *Network Operations Report Summer 2008*. Eurocontrol, December 2008.
- [15] D. Ross, "Game theory," in *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/fall2009/entries/game-theory/>, 2009.
- [16] A. Ranieri and L. Castelli, "A market mechanism to assign air traffic flow management slots," in *Proceedings of the VIII USA/Europe Air Traffic Management R&D seminar, Napa, California, USA*, June 2009.
- [17] Eurocontrol, *Basic CFMU Handbook - CFMU General & CFMU Systems*, 13th ed. Eurocontrol CFMU, 2009.
- [18] A. Cook, G. Tanner, V. Williams, and G. Meise, "Dynamic cost indexing - managing airline delay cost," *Air Transport management*, Elsevier Ltd., vol. Vol. 15, Issue 1, pp. pp. 26–35, January 2009.
- [19] N. Barnier, P. Brisset, and T. Rivire, "Slot allocation with constraint programming: Models and results," in *Proceedings of the IV USA/Europe Air Traffic Management R&D seminar, Santa Fe, NM, USA*, December 2001.
- [20] N. Barnier and C. Allignol, "4d-trajectory deconfliction through departure time adjustment," in *Proceedings of the VIII USA/Europe Air Traffic Management R&D seminar, Napa, California, USA*, June 2009.
- [21] R. Hoffman, M. O. Ball, and A. Mukherjee, "Ration-by-distance with equity guarantees: a new approach to ground delay program planning and control," in *Proceedings of the VII USA/Europe Air Traffic Management R&D seminar, Barcelona*, July 2007.
- [22] J. Rosenberger, E. Johnson, and G. Nemhauser, "A robust fleet-assignment model with hub isolation and short cycles," *Transportation Science*, vol. 38(3):357-369, pp. 357–369, 2004.
- [23] D. Klabjan, E. L. Johnson, and G. L. Nemhauser, "Airline crew scheduling with regularity," *Transportation Science*, vol. 35(4), pp. 359–374, 2001.
- [24] M. Athans, P. Falb, and R. T. Lacoss, "Time-, fuel-, and energy-optimal control of nonlinear norm-invariant systems," *IEEE Transactions on automatic control*, vol. 8, pp. 196–202, Jul 1963.
- [25] L. Delgado and X. Prats, "Fuel consumption assessment for speed reduction solutions during the cruise phase," in *Proceedings of the Conference on Air Traffic Management Economics, Belgrade*, September 2009.
- [26] Boeing, "Fuel conservation strategies: cost index explained," Boeing, Tech. Rep., 2007.
- [27] K. Virtanen, H. Ehtamo, T. Raivio, and R. P. Hmlinen, "Viato visual interactive aircraft trajectory optimization," *IEEE Transactions on systems, man, and cybernetics*, vol. vol. 29, pp. 409–421, 1999.
- [28] L. P. de Menezes Junior, "Metodos para a reduo de custos operacionais em empresas aeras," Ph.D. dissertation, Instituto Tecnologico de Aeronautica, So Jos dos Campos, 2006.
- [29] L. Delgado and X. Prats, "En-route speed reduction for the management of atfm delays," in *Proceedings of the 9th AIAA Aviation Technology, Integration, and Operations Conference (ATIO)*, Hilton Head, South Carolina, September 2009.
- [30] I. A. B. Wilson, "4-dimensional trajectories and automation connotations and lessons learned from past research," in *Proceedings of the Integrated Communications, Navigation and Surveillance Conference (ICNS)*, Herndon, Virginia (USA): IEEE, Apr 2007.
- [31] J. T. Betts and E. J. Cramer, "Application of direct transcription to commercial aircraft trajectory optimization," *Guidance, control and dynamics*, vol. 18, no. 1, pp. 151–159, January-February 1995.
- [32] Airbus, *A320, Flight Crew Operating Manual, Book 2, Flight Preparation*, Airbus, Ed. Airbus Training.
- [33] ———, *A320, Flight Crew Operating Manual, Book 3, Flight Operacions*, Airbus, Ed. Airbus Training.

High-fidelity Human-in-the-Loop Simulations as one Step Towards Remote Control of Regional Airports

A Preliminary Study

C. Möhlenbrink, M. Friedrich, A. Papenfuß, M. Rudolph, M. Schmidt, F. Morlang, N. Fürstenau

German Aerospace Center (DLR)
 Institute of Flight Guidance
 Christoph.Moehlenbrink@dlr.de

Abstract—The paper reports on an experimental work environment for simulating remote control of regional airports and initial results obtained by high fidelity human-in-the-loop simulations. At the Institute of Flight Guidance of the German Aerospace Centre a concept for remote control of regional airports was developed since 2002 and a corresponding experimental testbed realized, consisting of facilities for field testing at the Braunschweig research airport and a tower simulator extension for operational remote tower (RTO) simulations with controllers. Human-in-the-loop simulations were conducted, simulating Braunschweig airport to show the operational feasibility of the new working environment. Therefore two tower controllers handled traffic scenarios using a common 200-degree tower simulator, but also using the new work environment, the RTO (Remote Tower Operation)-Console. This setting allows a direct comparison of an evaluation of the RTO-Console and the tower simulator as work environment. Augmented vision aspects were implemented in the simulation runs at the RTO-Console. Moreover, a zoom camera with an automatic tracking function integrated in the work environment for remote control was evaluated. Subjective data from questionnaires and free interviews were gathered for each simulation run. Objective eye data were recorded for the simulation runs using the RTO-Console. The main result from the questionnaires depicts the work environment of the RTO-Console to be comparable with working in a tower simulator. The eye data show that most of the time (53%) the tower controller is looking at the area of interest in the simulated far view, which is in line with former work analyses. The results of the human-in-the-loop simulation suggest the feasibility of tower operation using the RTO-Console. For the operational deployment of remote control of small airports a stepwise validation using human-in-the-loop simulations is indispensable.

Keywords - Remote Tower, Controller working position, simulation, validation, eye-tracking

I. INTRODUCTION

The concept of remote control of small airports is a great challenge with regard to the implied changes in the controllers

work environment. At German regional airports air traffic control is provided by a controller team located in a tower with direct view on the airport. A concept for remote control of small airports must guarantee safe operations. One important question within this context relates to “which” information an air traffic controller needs to guarantee safe operations on the airport.

Research in the past did not focus on the work environment of tower control. According to Hein [1] research focused on radar controller positions due to the higher potential for implementing new technologies. Hagemann [2] cites that automation for tower controllers seems impossible due to the high number of unforeseeable events. He mentions for example events like runway inspections by staff of apron control, obstacles on the runway or rejected landings or takeoffs. Even though these events are not linked to the direct line of sight from the tower, it seems that it plays a major role in this context. The importance of the far view for tower control is also pointed out by Hilburn [3] and concluded from a literature review accomplished by Tavanti [4]. However the role of the far view for air traffic control at airports remains inexplicit.

It is conceivable that this lack of understanding is the reason why remote control of regional airports is not realized yet. From 2002–2004 a concept study „Virtual Tower“ was accomplished at DLR, that initialized research concerned with remote control of small airports. Within the DLR project RAPOr (Remote Airport Tower Operation research) first steps of the concept were realized [6][6, 7]. From the arguments in the last paragraph and a task analysis at Leipzig airport, supported by the project partner DFS (German Air Traffic Control), it was concluded that the far view is a crucial information source in the safety chain of air traffic control that cannot be neglected for remote control. Therefore the work environment for remote control of airports developed, involves a reconstruction of the far view by a live stream video panorama with 180° field-of-view (FOV) and an additional zoom camera. For demonstration and proof of the technical concept an experimental system was realized at research airport Braunschweig. For more details of the camera system and the configuration of the RTO-Demonstrator as working environment, see the references mentioned above.

After the demonstration of technical solutions for the live stream video panorama and the design of the RTO-Demonstrator an ongoing study is focusing on the operational aspects of the new work environment.

A. Aerodrome control: work environment, working positions, and tasks

The direct view on the airport is just one visual information source. Flight strips, RADAR and a weather display are also on the controller's disposal. Beside these visual information sources, the controller is using radio especially for communication with the pilots, ground radio and a telephone. These visual and auditory information sources and interaction devices are the main instruments for air traffic controllers on a regional airport [8]. For the aerodrome control of a regional airport there is no separation between tower and approach control. Usually a tower controller and coordinator work in a team together. Their tasks are the control and surveillance of the runway, taxiways and park areas and the surveillance of the whole aerodrome. The tower controller is in contact with the pilots via radio and responsible for the safe operations on the airport, while the coordinator or assistant is more concerned with planning the arriving and departing traffic and assists the tower controller with the protocol on the flight strips. The main difference of the tower control or aerodrome control in comparison to approach or radar control is the direct view out of the window. Although the role of the outside view is still not understood, it is one element in the safety chain for air traffic control to justify safe operations on the airport.

B Characteristics of regional airports and the consequences for air traffic control

Small and regional airports usually lack any electronic surveillance like ground radar. Another characteristic of regional airports is the high percentage of flights operating under Visual Flight Rules (VFR) conditions. Flights under Instrumental Flight Rules (IFR) in large parts are operated by commercial airline pilots. In contrast, flights operating under VFR conditions are mainly flown by private pilots and pilot trainees. These characteristics go in line with several consequences for air traffic control at regional airports. The VFR traffic can not be foreseen like the IFR traffic, and therefore demands flexible planning. The fact that pilots of VFR traffic are in general less experienced pilots has influences on the air traffic control service. It is likely that there are more deviations from the standardized controller-pilot radio communication and there is less confidence, that the pilots follow the commands from the tower correctly. This study aims at validation of the video panorama. It is assumed that the RTO-Console is able to represent the crucial changes for the work environment provoked by the RTO-Console.

C Information superimposition into the video panorama and automatic zoom camera tracking for controller assistance

The replacement of the outside view by a live video as suggested for a remote control working position offers new possibilities for controller assistance. This is interesting especially for poorly technically equipped regional airports. The live video offers possibilities like image processing algorithms to automatically detect aircraft or other moving objects [9]. Some research has been performed on augmented vision, i.e. superimposition of flight information on the far view within the real tower environment [10, 11, 12, and 13]. Augmentation is more easily realized within the video panorama because for the vehicle tracking case no alignment of augmenting information between distant moving vehicle and observer position is necessary [6, 7, 9]. Within the following simulation study, information superimposition with object tracking and semi-automatic zoom camera tracking will be integrated into the experimental remote tower environment.

D Realization of an operational concept for remote control in a high fidelity simulation environment

The realization of an operational concept for remote control of regional airports in a high-fidelity simulation environment is important for a stepwise validation. The RTO-Console with its four displays for the panorama view and an additional display for the zoom camera was therefore integrated into the tower simulation environment at DLR Braunschweig. The aim of the following study is to investigate the operational feasibility of an RTO work environment offering a video panorama of the airport and a zoom camera to one controller team (tower controller, coordinator). In addition to that a focus is set on the examination of new possibilities of controller assistance at the remote working position. The utility of a zoom camera with a semi-automatic tracking function and the utility of information superimposition will be investigated.

As methodological approach for validation of the novel work environment, working in a common 200-degree tower simulator is compared to working with the 180° RTO-Console, simulating airport Braunschweig. In the following human-in-the-loop simulations the four different conditions were realized that are now introduced.

1) Tower simulator

In the first simulation run one controller team works in a conventionally high-fidelity tower simulator, with a four-beamer-front projection to achieve sizes and proportions adequate to the look out of a tower window. No zoom camera is available to the controllers

2) RTO-Console Baseline

This experimental condition is for the straight comparison of the tower simulator and the RTO-Console. A controller team is working at the RTO-Console, representing the outside view reduced to 30° via 4-beamer rear projections. A zoom camera was available for the tower controller that was controlled by a virtual joystick using a touchpad input device.

3) RTO-Console and image processing

Compared to condition 2, image processing algorithms are realized for the detection of moving objects. Moving objects are visualized by color frames in the outside view. It is further possible for the tower controller to direct the zoom camera on one moving object and start the automatic tracking of that object by a touch pen.

4) RTO-Console and data fusion

It is assumed, that in the near future also regional airports are able to receive and process position data via MODE-S transponder or GPS-ADS-B. These data involve position data, but also the callsign of the aircraft. In this condition, not only a color frame, but the callsign is displayed next to the aircraft in the outside view. Moreover the transponder data are used for the semi-automatic tracking function, like in condition 3.

II. METHOD

A. Subjects

Two tower controllers from Braunschweig tower participated in the high-fidelity simulation study. Both participants operate as tower controllers since more than 25 years and are working in Braunschweig since more than 20

years. Their mean age was 53 years. They were announced an expense allowance.

B. Experimental Design

A 2*3 factorial quasi-experimental design was tested with the factors weather (good visibility, bad visibility) and working position variants (tower simulation, RTO-Console baseline, RTO-Console + augmented vision). The weather condition was varied between day one and two. On day one the augmented vision aspects were realized using image processing. On day two the callsign (data fusion) was displayed. The experimental design is depicted in table 1.

TABLE I. EXPERIMENTAL DESIGN

	Simulation day 1 (good visibility)	Simulation day 2 (bad visibility)
Baseline	Tower simulator	Tower simulator
Baseline	RTO-Console (no zoom function)	RTO-Console (no zoom function)
Augmented vision	RTO-Console › image processing › zoom camera + automatic tracking	RTO-Console › data fusion (callsign) › zoom camera + automatic tracking

C. Simulation Setup

The RTO-Demonstrator was integrated into the ATM-simulation environment at DLR in Braunschweig. For the tower simulator a 4-beamer-front-projection on a 200° spherical surface was realized, imitating the view out of the window without shrinking of sizes and proportions. The RTO-Console was equipped with a 4-beamer rear projection on 1200*1600px resolution, 30' displays. The video on RTO-Console was brighter, had a better contrast, saturation and sharpness in comparison to the tower simulator. The simulation setup of both, the tower and RTO simulation comprises two controller working positions (tower controller, coordinator), two pseudo pilot working positions and a supervisory working position. The integration of these five working positions into the ATS with the Simulation Server, image generator and environment is depicted in figure 1.

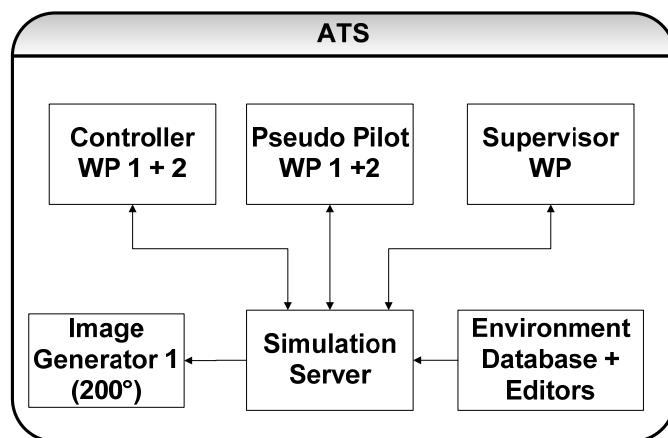


Figure 1. Schematic simulation setup

D. Working positions

In the following the five working positions are introduced, before a description of the generated traffic scenario is given.

The two controller positions (tower controller and coordinator) are equipped with a generic RADAR application, a weather display, flight strips, radio and the far view. Within the tower simulator two RADAR applications were used, because of the large distance from the coordinator working position to the RADAR placed in front of the tower controller position.

At the RTO-Console there is an additional control display in front of the tower controller for the zoom camera, but just one RADAR right in front of the coordinator working position. The configuration of working positions are depicted in figures 2 and 3.



Figure 2. Design of working positions in the tower simulator compasses [1] RADAR, [2] weather display, [3] flight strips, [4] far view.

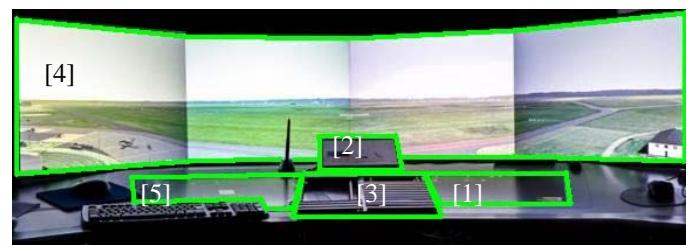


Figure 3. Design of working positions of the RTO-Console [1] RADAR, [2] weather display, [3] flight strips, [4] far view (videopanorama), [5] controldisplay (zoom camera)

At the two pseudo pilot working positions, located in a separate room, two trained pseudo pilots control the aircrafts by a commando pad and mouse clicks, according to the radio advices from the tower controller. They are responsible for heading, speed and altitude advisories and the clearances of all aircraft in the simulation.

The supervisor is responsible for the start-, freezing- and stop- commands in the simulation.

The traffic scenarios (30 min) differed in actual callsigns of the aircraft, so that there was no learning effect over the different simulation runs, but all were designed with a similar demand and overall traffic mix of 60% VFR-traffic. The simulation scenarios always allowed VFR-flight operations even with weather variations.

E. Data acquisition

1) Questionnaires and interview

After each simulation run, both controllers had to fill out a questionnaire concerning the usability and acceptance of the whole system. Therefore questions from the EUROCONTROL SHAPE questionnaires were used, that were developed to access the influence of automation on workload, situational awareness, team work and trust [14]: Parts of the SHAPE-AIM (assessing the impact of automation on mental workload), parts of the SHAPE-SATI (system automation trust index) and parts of the SHAPE-SASHA (situational awareness for SHAPE) were used. Every item was rated on a seven-point Likert scale. Further questions from the system usability scale (SUS) were used to access subjective rating concerning the usability of the RTO-Console on a five-point Likert scale [15].

Afterwards the free interview was used, to record the feedback of the controllers concerning the RTO working environment and the novel system components (zoom camera, touch pen and tracking function).

2) Eye-tracking

During the simulation runs at the RTO-Console eye tracking data were recorded. The eye data are valuable objective data on the information acquisition of the tower controller and the coordinator. However, eye-data recording in such complex environments like the tower simulator or RTO-Console is not trivial. Hence controllers look at quite a lot different information sources and might stand up from their chairs to have a better look at specific areas on the airport. The head-mounted system iView-X from company SMI has the disadvantage that it comes in its original form in combination with a magnetic head-tracker (polhemus) for data based analyses [16]. Such head-trackers implicate about three disadvantages: They need a metal free environment, (2) a maximal distance of 80 cm between sensor (head) and receiver (polhemus), and (3) can just track the head of one person. Thus within the project RAiCe (Remote Airport Traffic Control Center) the coupling of an optical tracking system from A.R.-tracking with the iView-X-System from SMI was realized [17]. The system uses two infrared cameras, to track the head using "targets" and can overcome all three disadvantages.

F. Data analyses

1) Data analyses of questionnaires

Only a subset of the questions of the SHAPE questionnaires was used in the study. 15 questions for mental workload (AIM), 7 for system trust (SATI) and 4 for situational awareness (SASHA) were selected. The questions relevant for an evaluation of the four different test conditions. Here, mean values are calculated for the items of each questionnaire for a relative comparison. Thereby it is respected that several items had to be inverted in their values.

For the contrast of the different experimental conditions the sum scores of the different controller positions (tower controller, coordinator) and simulation days (one, two) are aggregated (compare table 1). According to these aggregated data the different working conditions (1) tower simulation, (2) RTO-Console Baseline and (3) RTO-Console (Augmented Vision) are compared on descriptive level. Inferential statistics are denied for the sample of two controllers.

The comments of the free interview will be summarized according to four categories: (1) design and arrangement, (2) augmented vision aspects, (3) other suggestions, (4) general comments.

The category "design and arrangement" will consider the arrangement of the instruments at the RTO-Console. Under "augmented vision aspects" all comments concerning the superimposition of information and the usability of the zoom camera will be summarized. "Other suggestions" will report controller comments in respect to the design, arrangement and augmented vision aspects and the last category "general comments" is for remaining remarks.

2) Analyses of eye-tracking data

The analysis of eye-tracking concerns the dwell times for the different information sources as an index for the visual attention distribution.

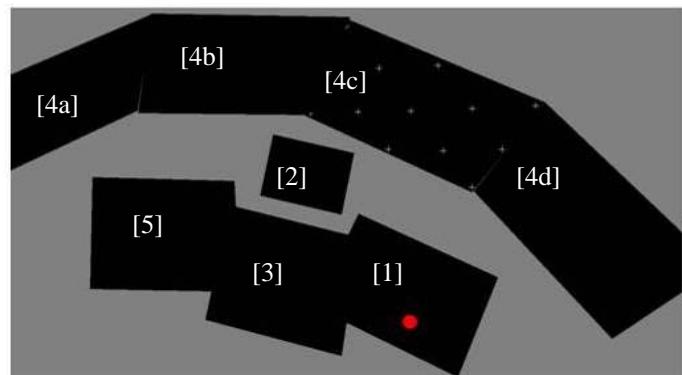


Figure 4. 3d-Modell of the information sources (planes) of the RTO-Console as basis for data driven analysis of eye-tracking data

The present analyses will focus on the differences of this index between the tower controller and the coordinator. For these data analyses it is essential to generate a 3d-model of the RTO-Console, which includes the positions of all visual information sources (planes). It can be seen in figure 4. On plane 4c there are 13 white crosses used for the calibration of the eye tracking system. All four displays of the panorama will be defined as four separate planes (4a-4d). Due to the layout of airport Braunschweig the four displays can be assigned to four different functional areas of the airport that are mainly represented on a specific plane.

- 4a) Departure - due to the runway direction 26 used in all simulation runs
- 4b) Park area – this display includes the park areas
- 4c) Runway – on this display (sometimes extended into 4b)) the touch down of aircraft takes place
- 4d) Approach- due to the runway direction 26 used in all simulation runs

Following the eye data recording based on the 3d-model, the eye-tracking analyzer (EyeTA) is used. The EyeTA is a tool that was developed at the DLR Institute of Flight Guidance, to be able to analyze large eye-tracking data sets even for a large sample size. For calculation of fixations the algorithms suggested by Salvucci and Goldberg were implemented [18]. The user interface is shown in figure 5. The tool enables to load multiple eye-data sessions, to define parameters like a threshold value to differentiate between fixations and saccades or minimum fixation duration. It offers the possibility of half-

automated data analysis even for long eye-data sessions and studies with large sample sizes. Therefore it is a strong tool for the analysis of eye data for the full RTO study.

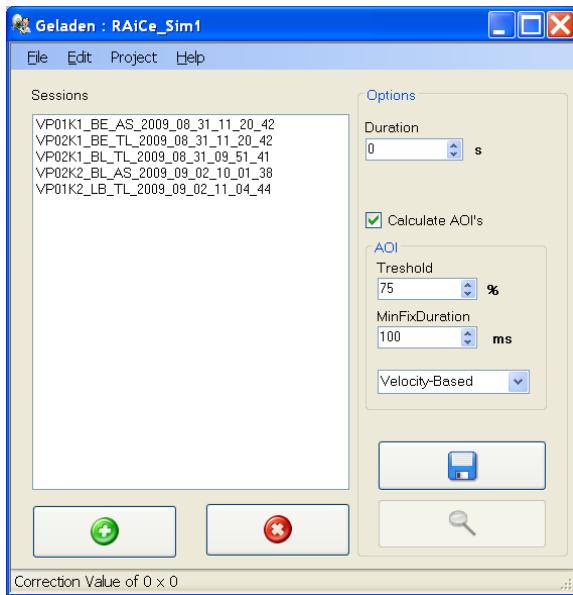


Figure 5. User interface of the eye-tracking analyzer

III. RESULTS

The present section is subdivided into subsection 3.1 reporting the results of the questionnaires, subsection 3.2 which addresses data of the free interview, and subsection 3.3 presenting the eye-tracking results.

A. Questionnaires

The results of the questionnaires are represented in figure 6. For mental workload (AIM) a mean value of 29,8 (STD=6,13) is found for the run tower simulator, 30,3 (3,86) for the RTC-Console (baseline) and with 25,8 (4,92) the least value is found for RTC-Console (with augmented vision). For the system trust in automation (SATI) the mean sum values are 39,8 (2,5); 39,3 (2,5) and 41,0 (4,1) and for the situational awareness (SASHA) the mean sum scores are 6,5 (1,9); 6,3 (3,6) and 5,5 (1,7) respectively.

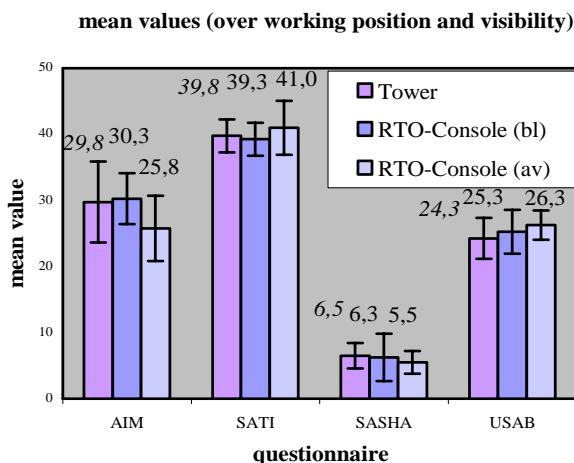


Figure 6. Sum scores of the questionnaires aggregated over working positions and visibility

Though the differences are very small, the RTC-Console (augmented vision) shows the highest system trust in automation index. Finally for an evaluation of the usability the USAB shows the smallest value for the tower simulator with 24,3 (3,1). The evaluation of the RTC-Console (baseline) ended up with a value of 25,3 (3,3). The highest value was given to the RTC-Console (augmented vision) with 26,3 (2,2) points. Cohen's d effect sizes were calculated and showed a medium effect size for mental workload (AIM) for the comparison tower simulation and RTO-Console (augmented vision): $d=.72$. There is less negative influence on mental workload for the augmented vision condition. Another medium effect size was found for the system usability ($d=.74$) reporting a better usability for the augmented vision condition in contrast to the tower simulator condition. Small effect sizes became apparent for the system trust in automation and the situational awareness. In the RTO-Console with augmented vision condition there are less negative effects on situational awareness ($d=.55$) and there is a better system trust ($d=.37$) than for the tower simulation condition. For a comparison between the tower simulator and the RTO (baseline) condition no effects were found, but one. A small effect size ($d=.31$) reported a better usability of the RTO-Console (baseline).

B. Free interview

The results of the free interview will be summarized in four predefined categories.

1) Design and arrangement

For the RTO-Console both controllers mentioned at the end of simulation day one to place the weather display in the centre of the RTO-Console. On left side, it was hardly seen by the coordinator and the tower controller complained the long distance between weather display and RADAR. For the RADAR both controllers ask for a steeper angle to facilitate the view on the RADAR from the tower position.

2) Augmented vision aspects

The augmentation of moving objects by a colour frame was rated as unnecessary by the tower controller. He mentioned to know about all aircraft positions and that he can't see the assistance value. In contrast the coordinator rated the frames more positive as they can help especially the coordinator to update his picture about the traffic on the airport, after longer planning tasks focusing on the RADAR and the flight strips. The superimposition of the callsign into the far view as it was realized on the second day of simulation was rated good and valuable. It helps not to confuse the callsigns when aircraft line up for take off. It was also mentioned that the callsign can be valuable, if it is placed right next to the aircraft, to narrow the field of search for arriving traffic at the horizon. The ability to use a touch pen to activate an automatic tracking of the zoom camera for arriving traffic was rated as a positive characteristic of the RTO-System. One controller mentioned that within the daily operations at a tower, the controller is using the binoculars quite often, to track arriving aircraft.

3) Other suggestions

The controller recommended activating the callsigns only for aircraft, which are in contact with the tower controller via radio. After an aircraft has reached its parking position, or an aircraft has not yet asked for taxiing the callsign is of no interest for the controller team.

4) General comments

One controller mentioned that he preferred working at the RTO-Console compared to the tower simulator. He realized that the resolution of the simulation and the brightness was much more comfortable at the RTC-Console. Even the touch-and-go manoeuvres were clearly visible at the RTC-Console. Regarding the question what the most important information sources are, it was mentioned, that for the tower controller the most relevant information source is the outside view and radio, in contrast to the coordinator who uses the RADAR and flight strips most often. These subjective rating will be compared with the objective eye data in the discussion section.

C. Eye-tracking data

The recording of the eye-tracking was successful for the tower controller in the simulation run RTC-Console (baseline, day1) and for both controllers in the simulation run RTC-Console (image processing, day1). On the second day only one eye-tracking system was available and therefore recording was conducted for the coordinator in the baseline condition and the tower controller in the datafusion condition. An overview of the eye data sets is given in table 2.

TABLE II. OVERVIEW ABOUT PRESENT EYE DATA

Simulation runs with recording of eye data	tower controller	coordinator
day1: RTO-Console (baseline)	successful recording	recording not successful
day1: RTO-Console (image processing)	successful recording	successful recording
day2: RTO-Console (baseline)	no recording	successful recording
day2: RTO-Console (data fusion, callsign)	successful recording	no recording

As a first index the fixation and macrofixations were calculated for each simulation run. It arises, that there are comparable numbers of fixations and macrofixations for all simulation runs with successful data recording. The minimal number of fixation appears for the tower controller in the RTO-Console baseline condition with 3035 fixations. Most fixations are found for the tower controller in the RTO-Console datafusion condition, displaying the callsign. The macrofixation also vary within a narrow range between 605 and 705. The relative frequencies of fixational dwell times for single AOI's as well as AOI's aggregated for single controllers exhibit clustered Poisson statistics, i.e. somewhat steeper-than-exponential probability density functions.

No eye data were collected during the simulation runs at the tower simulator. A comparison of the different experimental conditions at the RTO-Console (BL, BE, CA) did not show significant differences. In the following a comparison of the controller positions (tower controller, coordinator) will be considered as input for the relevance of instruments for each working position at the RTO-Console.

The five data sets (coordinator (n=2), tower controller (n=3)) are used to compare the mean percental dwell times of the coordinator and the tower controller on the different information sources. For a graphical representation of these data regard figure 7. While the tower controller directs his visual attention to the video panorama for 53%, the coordinator spends just 34 % on the outside view. He is fixating the flight strips 34% of the time and the RADAR 28% of the time. The

tower controller is just looking at the flight strips 13% and on the RADAR 17% of the time.

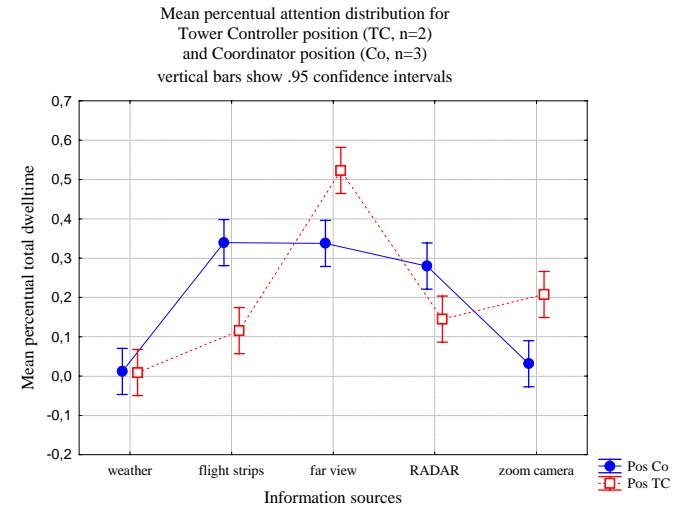


Figure 7. Mean percentual distribution of the visual attention on the different information sources, according to the working positions.

Further it is calculated which areas of the far view are of interest in respect to the two working positions. The data show that the tower controller and coordinator both spend 21% of the time looking in the approach area. The tower controller spends 24% looking at the departure area, while the coordinator is looking into this area only 8% of the time. Both controller positions spend about 5% with looking on the runway and 7% and 5% respectively on the Apron area. The results are summarized in figure 8.

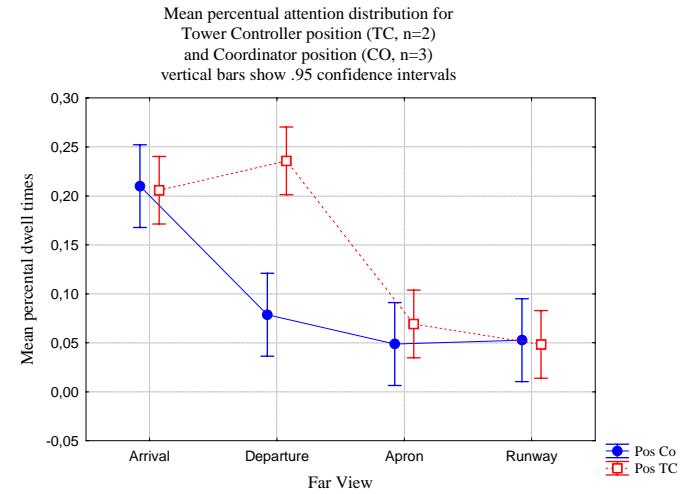


Figure 8. Mean percentual distribution of the visual attention on the different information sources, according to the working positions.

IV. DISCUSSION

A. Subjective Evaluation of the RTO-Console (questionnaire)

The subjective rating of the two controllers indicated that working at the experimental RTO-Console there is no negative effect on the situational awareness (SASHA) no increase of workload (AIM), no negative effect on the system trust in automation index (SATI) and no decrease in ratings about the system usability (SUS). It has to be mentioned that in this

preliminary study it was not possible to control for sequence effects and for strong conclusions the results of the full study have to be considered.

The strongest validation for an RTO-Console is a comparison of a work sample in a real physical tower with a RTO-Console simulation. This was not possible within the study, so a tower simulation was used to compensate this aspect. The comparison of the results working in the tower simulator with working at the RTO-Console for validation is accepted insofar, as human-in-the-loop simulations in a tower simulator are an accepted method to draw conclusions about the working conditions in a real physical tower.

To the knowledge of the authors, this is the first simulation study for the validation of an RTO work environment. Even though the results from the questionnaire have just descriptive character ($n=2$) the ratings concerning situation awareness, work load and usability are important to ensure, that the changes of the design of the work environment (RTO-Console) have no negative effect on the controllers' operation to do air traffic control service.

B. Free interview

The feedback concerning the system arrangement was positive and the minor changes of the weather display and RADAR do not need further discussions.

The feedback concerning augmented vision aspects was in general positive. Superimposition by frames that depict detected moving objects, were seen as helpful for the coordinator. For superimposition of the callsigns the controllers mentioned two concrete positive aspects, reducing confusion of callsigns and narrowing the search of arriving traffic. In general the superimposition of information is motivated by the surveillance task of the tower controller. According to Hilburn head-down times are the major contributors to the risk of missing critical events [3]. In his study head-down time defines the time, when the controller is not looking out of the tower window. The effectiveness of information superimposition has been researched extensively for head up displays (HUDs) in the cockpit [19]. Research has shown, that the representation of additional information can also have negative effects on pilots' attention. The superimposed information can become a distractor or elicits effects like cognitive tunneling. It is the task of future research to learn, if minimizing head-down times by superimposition will in fact minimize the detection of critical events and if negative side effects can be excluded.

The use of a zoom camera for automatic tracking of landing and starting aircraft was also rated positive and easy to use. The feedback pointed out two aspects for the question, which information is needed by the tower controller: He is looking for "gear down" of arriving traffic and if an arriving aircraft definitely left the runway after touch down. This information is checked visually, according to today's operation. A display, representing a zoom video tracking an arriving and landing aircraft, can be a valuable assistance tool for the controller to get this information.

Other suggestions of the controllers pointed out, that a classification of information into relevant and irrelevant information is important for augmented vision aspects. If a tower controller's work environment will include a video panorama the superimposition of information will be easy. However in the simulation runs the controllers were not interested in callsigns not in contact via radio with the controller. Research must set more effort to identify the task

relevant information, not to display too much information in the video panorama.

The final feedback one controller mentioned the difference that the tower simulator has a worse resolution and less brightness compared to the RTO-Console. It is conceivable that this fact is an explanation for the better ratings on the SUS in the RTO conditions.

C. Objective Eye Data

The reduced eye data set does not allow a comparison of working at the RTO-Console and working in the tower simulator. The analyses of the eye gaze position within the new working environment in this study focuses on the visual attention distribution on displays and instruments and on a comparison of both controllers. These results are interesting for the arrangement of the instruments at the remote tower console.

The eye data show, that the tower controller spends most visual attention on the panorama, which resembles the view out of the window. These values are in line with reported results in the literature [20]. What is not found in the literature is the comparison of the visual attention contrasting the tower controller and the coordinator working positions. The objective eye data show that the most used information source is the view on the panorama video. The coordinator is looking most of the time on the RADAR and on the video panorama followed by the flight strips. The difference between the working positions can be explained with the different tasks the controllers have to conduct. It is interesting to note that they accord with the comments of the controllers', that the most important information source for the tower controller working position is the view out of the window and radio, while for the coordinator the RADAR and the flight strips are much more of interest.

A more detailed analysis of the far view shows a big difference between the two controller positions on the departure display. On this display it is possible to see, if a landing aircraft left the runway and if departure traffic is taking the flight routes commanded by the controller. The data support that this information must be confirmed by the tower controller. It is not the task of the coordinator to verify this information by looking out of the window.

V. SUMMARY

Overall, working at the RTO-Console and the traffic scenarios were seen as realistic. The results of the questionnaires show, that working in the simulation condition RTO-Console with augmented vision got the most positive ratings. The results on situational awareness (SASHA) show the least mean value, with a medium effect size. This indicates a better rating for situational awareness in the RTO-Console with augmented vision condition compared to the tower simulation condition. The responses for mental workload (AIM) also show the lowest mean value (medium effect size) indicating the lowest workload, respectively. The system trust in automation index (SATI) and the mean values of the system usability scale reveal the highest scores for condition RTO-Console with augmented vision. These values assigning the RTO-Console the best scores for trust and usability only show small effect sizes.

However all these data report, that we cannot find a break-in for any of the scales like mental workload, situation awareness, system trust and usability, for the RTO conditions,

representing a solution for remote control of airports. It is suggested that it is the idea of the reconstruction of the far view that causes good values especially for the psychological constructs like mental workload and situational awareness.

Limitations of the comparison between the common tower simulator and the RTO-Console and the problems for expansive conclusions from these data due to sequence effects were issued in the discussion section.

Within this preliminary study, the free interview provides an indication that the design of the RTO-Console as it was realized enables a controller team to handle the traffic of a regional airport successfully. The comments about the weather display and the angle of the RADAR display are not critical for the general operational concept, but aim at optimization of the instrument arrangement at the RTO-Console.

The superimposition of information was accepted by the controllers and the zoom camera was appreciated as replacement of binoculars. It is useful to verify important information e.g. that a landing aircraft left the runway.

The eye data indicate which information is important for the controllers, and how long he is looking on different information sources. This objective method is valuable, especially for a better understanding of the role of the far view. The data indicate which information is acquired while controllers fulfill their air traffic control service.

VI. OUTLOOK

This paper introduced a preliminary study with a sample size of 2 controllers of airport Braunschweig. The sample size will be enlarged so that the full study will include data of 12 tower controllers. This will allow to calculate inferential statistics and to control for sequence effects and interpersonal differences. Ongoing work addresses application of advanced statistical analysis to eye movement, like moving average windows allowing for event-based data analysis.

ACKNOWLEDGMENT

We are indebted to controllers from Braunschweig airport for participating in this simulation study and their feedback to the operational issues relevant for the simulation setup.

REFERENCES

- [1] Hein, M. (2003). Optimierungspotentiale der Schnittstelle zwischen zentraler Vorfeld- und Platzkontrolle an deutschen Verkehrsflughäfen. PhD, Fakultät für Humanwissenschaften und Theologie, Universität Dortmund.
- [2] Hagemann, T. (2000). Belastung, Beanspruchung und Vigilanz in der Flugsicherung unter besonderer Berücksichtigung der Towerlotsentätigkeit, PhD, Peter Lang.
- [3] Hilburn, B. (2004) Head-down Time in Aerodrome Operations: A Scope Study, Center for Human Performance Research (CHPR). The Hague, the Netherlands
- [4] Tavanti, M. (2006) Control Tower Operation: A literature review of task analysis studies. EEC Note No 10/06.
- [5] Fürstenau, N., Rudolph, M., Schmidt, M., Werther, B. (2004): Virtual Tower, in: " Wettbewerb der Visionen 2001 – 2004", Hrsg. Deutsches Zentrum für Luft-und Raumfahrt (2004) pp.16 – 21
- [6] Fürstenau, N., Schmidt, M., Rudolph, M., Möhlenbrink, C., Werther, B. (2007): Augmented Vision Videopanorama System for Remote Tower Operation: Initial Validation, Proc. 6th Eurocontrol Innovative Research Workshop, Bretigny, 4.-6. 12. 2007, 125-132
- [7] Schmidt, M., Rudolph, M., Werther, B., Fürstenau, N. (2007): Development of an Augmented Vision Videopanorama Human-Machine

Interface for Remote Airport Tower Operation, Proc. HCII2007 Beijing, Springer Lecture Notes Computer Science 4558 (2007) 1119-1128

- [8] Papenfuß, A & Möhlenbrink, C (2009) *Kognitive Arbeitsanalyse Lotsenarbeitsplatz*, 112-2009/20, DLR, Braunschweig.
- [9] Fürstenau, N., Schmidt, M., Rudolph, M., Möhlenbrink, C. Halle W., (2008). Augmented vision videopanorama system for remote airport tower operation. In: Proc. ICAS 26th Int. Congress of the Aeronautical Sciences. I. Grant (Ed.), Anchorage, ISBN 0-9533991-9-2 paper 93
- [10] Fürstenau, N., M. Rudolph, M. Schmidt, B. Lorenz, T. Albrecht, (2004). On the use of transparent rear projection screens to reduce head-down time in the air-traffic control tower, Proc. Human Performance, Situation Awareness and Automation Technology (HAPSA II), Daytona Beach , Lawrence Erlbaum Publishers Inc., 195–200
- [11] Schmidt, M., Rudolph, M., Werther, B., Fürstenau, N. (2006). Remote Airport Tower Operation with Augmented Vision Video Panorama HMI. In: Proc. 2nd Int. Conf. Research in Air Transportation ICRAT, Belgrade, Serbia , 221-230
- [12] Ellis, S., "Towards determination of visual requirements for augmented reality displays and virtual environments for the airport tower". Proc. NATO workshop on Virtual Media for the Military, West Point /N.Y., HFM-121/RTG 042 HFM-136, (2006) 31-1-31-9
- [13] Peterson, S., E.Pinska, (2006) Human Performance with simulated Collimation in Transparent Projection Screens. Proc. 2nd Int. Conf. Res. in Air Transportation, Belgrade, 231-237.
- [14] Eurocontrol (2008) The SHAPE Questionnaires. Verfügbar unter: http://www.eurocontrol.int/humanfactors/public/standard_page/SHAPE_Questionnaires.html
- [15] Brooke, J. (1996) SUS: a "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester & A. L. McClelland (eds.) Usability Evaluation in Industry. London: Taylor and Francis.
- [16] SensoMotoricInstruments (2007). SMI Eye & Gaze Tracking Systems. Verfügbar unter: <http://www.smivision.com> [24. Oktober 2009].
- [17] Advanced Realtime Tracking (2009) Eye-tracking. Verfügbar unter: <http://www.ar-tracking.de/Eyetracking.234+B6Jkw9.0.html> [24. Oktober 2009]
- [18] Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In Proc: Eye Tracking Research and Applications Symposium , 71-78. New York: ACM Press.
- [19] Crawford, J. & Neal, A. (2006). A review of the perceptual and cognitive issues associated with the use of Head-Up Displays in commercial aviation. International Journal of Aviation Psychology, 16 (1) 1-19.
- [20] Pinska, E. (2007): Warsaw Tower Observation. EEC-Note No. 02/07

Investigating String Stability of a Time-History Control Law for Interval Management

Lesley A. Weitz
The MITRE Corporation
7515 Colshire Drive
McLean, Virginia 22102-7508, U.S.A.
Email: lweitz@mitre.org

Abstract—Interval Management is a concept being developed as a part of the Next Generation Air Transportation System (NextGen) and Single European Sky ATM Research (SESAR). Within the FAA, standards and concept development are being sponsored by the Surveillance and Broadcast Services (SBS) office. The objective of Interval Management is to achieve a more precise spacing interval between a spacing aircraft and an assigned target aircraft. Speed commands, calculated by ADS-B-based avionics equipment, are implemented by the spacing aircraft in order to achieve a desired spacing interval relative to its target aircraft. In some Interval Management operations, it is expected that a string of aircraft will be formed, where each aircraft is spacing relative to its target, or preceding, aircraft. In the design of a speed control law to perform Interval Management operations, one must not only examine the performance and stability of one aircraft relative to another, but also the performance and stability of the entire string. String behavior fundamentally affects the potential operational practicality of successfully implementing Interval Management in certain operational environments. This paper presents a simplified, closed-form string stability analysis for a time-history speed control law, which has been proposed for Interval Management. Simulation results are shown to validate the closed-form analysis and are used to evaluate string behavior and system performance for an approach-spacing operation.

I. INTRODUCTION

Interval Management (IM) is a concept being developed jointly by US and European organizations to support the goals and objectives of the Next Generation Air Transportation System (NextGen) and Single European Sky ATM Research (SESAR), respectively. In the US, the FAA's Surveillance and Broadcast Services (SBS) office is sponsoring a significant effort to develop IM. The advent of ADS-B technology is leading to the development of cockpit-based avionics that take advantage of increased situational awareness available to the flight crew, and IM is one of a number of efficiency- and safety-enhancing concepts that are enabled by ADS-B. The goal of IM is to achieve precise inter-aircraft spacing as a means to increase the efficiency of a variety of operations. In IM operations, avionics equipment provides the flight crew of the spacing aircraft with speed commands that will achieve and maintain a desired spacing interval relative to one or more target aircraft. Approach spacing is considered to be an initial application of IM, where the operational goal is to increase the arrival capacity of busy airports by more precisely spacing

aircraft at the final approach fix (FAF) or runway threshold; in addition, controller workload will likely be reduced.

Previous research has demonstrated the benefits of delegating spacing or separation responsibility to the flight crew during approach operations in order to achieve more precise inter-aircraft spacing and reduce controller workload. Researchers have investigated using a cockpit display of traffic information (CDTI) to display relevant traffic information to the flight crew that enables self-separation without controller intervention [1], [2]. In addition, the large variations in low-noise and fuel-efficient approach trajectories have led to research into pilot tools and procedures that allow the more efficient trajectories without sacrificing capacity [3]–[5]. References [4] and [5], in particular, present a flap-scheduling tool designed to aid the flight crew in maintaining separation on approach and achieving a required time of arrival (RTA) at the runway, respectively.

In contrast to the previously mentioned research, IM operations will use a speed control law to calculate speed commands for the spacing aircraft to achieve and maintain a desired spacing interval. Commanded speeds are a function of the state information transmitted by ADS-B from one or more target aircraft. Researchers at NASA Langley Research Center have developed a speed control law for terminal area spacing, which uses estimated times of arrival (ETAs) at the runway threshold for both the spacing and target aircraft [6]–[9]. This is a trajectory-based speed control law, which assumes some knowledge of the planned arrival trajectories for the spacing and target aircraft in order to calculate the ETA. One advantage of NASA's trajectory-based control law is that aircraft on different routes can still be spaced precisely at the runway threshold. EUROCONTROL has developed and analyzed a time-history control law, which is sometimes referred to as a constant-time-delay control law, for spacing aircraft directly to a merge point after which the aircraft are on a common route. The time-history control law is designed to minimize the longitudinal, or along-track, spacing error between the spacing aircraft's position and the target aircraft's position τ seconds in the past, where τ is the desired spacing interval [10]–[12]. The primary advantage of EUROCONTROL's time-history control law is that information about the planned trajectories for the spacing and target aircraft is not required. Both of these speed control laws have had extensive performance testing

in simulation environments. Additionally, a variant of EUROCONTROL's time-history control law has been implemented and flight tested by United Parcel Service (UPS) [13].

In future approach-spacing IM operations, it is expected that a string of aircraft will be formed, where each aircraft implements speed changes in order to achieve the desired spacing relative to its immediately preceding target aircraft. In this case, the string stability of the selected speed control law must be investigated. String stability describes how spacing errors are propagated through the aircraft string as a result of disturbances or perturbations to target aircraft speeds. With a goal of developing IM equipment for near-term national airspace system (NAS) improvements, a greater understanding of string stability analysis and its implications for actual operation are needed.

The objective of this paper is to present string stability analysis for a time-history control law. Closed-form analysis for the string stability of a time-history control law is presented assuming a simplified aircraft model. Results from this analysis are related to well known results from the string stability literature. The closed-form analysis is intended to provide insight into string behavior for this control law and is the main contribution of the paper. Simulation results for a nonlinear aircraft model are used to validate the closed-form analysis and to relate that analysis to a realistic IM operation. Another technique for improving string behavior is explored using the simulation environment.

The paper is organized as follows. In Section II, additional background on the time-history control law and string stability is presented. In Section III, the closed-form string stability analysis is developed for a simplified aircraft model. Simulation results are shown in Section IV, and lastly, conclusions are presented in Section V.

II. BACKGROUND

A. Time-History Speed Control Law

The time-history speed control law presented here is considered for in-trail flight only, where the spacing and target aircraft are on a common route. The spacing error is defined as the difference in time between when the target and spacing aircraft cross a common point on the route and the desired spacing interval.

$$\text{spacing error}(p) \equiv (t_i - t_{i+1}) + \tau \quad (1)$$

Here, p refers to a common point on the route; t_i and t_{i+1} are the times when the target and spacing aircraft cross point p , respectively; and, τ is the desired (time-based) spacing interval.

A range error is used in place of the spacing error in the time-history control law. The range error is defined as the difference in the longitudinal position of the spacing aircraft at time t and the longitudinal position of the target aircraft at time $t - \tau$. The positions of the target and spacing aircraft are denoted as x_i and x_{i+1} , respectively.

$$\text{range error} \equiv e_i(t) = x_i(t - \tau) - x_{i+1}(t) \quad (2)$$

The commanded speed is a function of the range error term and the ground speed of the target aircraft at time $t - \tau$. The speed command to the spacing aircraft, v_{i+1}^c , is shown below, where the control gain k is a constant value.

$$v_{i+1}^c(t) = v_i(t - \tau) + k e_i(t) \quad (3)$$

Other implementations of a time-history control law have used gain scheduling to achieve more aggressive correction of range errors as the spacing aircraft gets closer to an endpoint [10].

B. String Stability

The time-history speed control law is designed to drive range errors, and equivalently spacing errors, to zero, which is indicative of a locally stable control law. However, string stability analysis, which reveals overall system performance, must also be considered for the case when a string of aircraft has been coupled through the speed control law. String stability deals with how spacing errors are propagated through the aircraft string due to changes in a target aircraft's speed.¹ A string-stable control law means that spacing errors between adjacent aircraft do not grow or amplify along the string of aircraft. In string-stable systems, spacing errors are attenuated along the string, and thus, aircraft far downstream from the disturbance will not detect the upstream disturbance.

The concept of string stability has been extensively studied for automated highway systems with vehicle strings of infinite length [14]–[16]. In the references listed here, a frequency-response approach is applied to determine whether the control laws in question yield a string-stable system. Using the spacing error definitions and control inputs, the spacing error transfer function that relates adjacent spacing errors is formed. The magnitude, or gain, of the spacing error transfer function determines whether spacing errors will propagate along the aircraft string.

In the literature, constant-distance control laws, which aim to maintain a constant distance relative to the immediately preceding vehicle only, are shown to yield a string-unstable system [14]. String stability can be gained by spacing relative to both the immediately preceding vehicle and the first vehicle in the string [14], [17], [18]. In contrast, a constant-time-headway control law, or a cruise control law, has been shown to be weakly string stable [15].²

Researchers at EUROCONTROL have investigated spacing-error propagation along a string for the time-history control law in a simulation environment. Results show that an additional parameter in the control law, the spacing anticipation time, τ_{sa} , can lead to an attenuation of spacing errors along the aircraft string [19]. The modified control law is shown

¹String stability is evident on busy highways when drivers speed up and slow down to maintain a desired distance relative to the preceding vehicle; i.e., the vehicles have a sinusoidal speed profile. The sinusoidal speeds have a greater impact on vehicles further along the string.

²In weakly string stable systems, the errors are neither amplified nor attenuated along the string for certain disturbances.

below; note that the spacing anticipation time is only applied to the target aircraft's ground speed.

$$v_{i+1}^c = v_i(t - \tau + \tau_{sa}) + k [x_i(t - \tau) - x_{i+1}(t)] \quad (4)$$

In the next section, the closed-form string stability analysis is presented for the time-history control law in (3). A similar closed-form solution could not be found for the control law in (4); however, numerical results are generated to provide insight on string stability for this modified control law.

III. CLOSED-FORM STRING STABILITY ANALYSIS

A. Transfer Function Derivation

To facilitate the closed-form string stability analysis, a simplified aircraft model is chosen. A single-degree-of-freedom, double-integrator model represents longitudinal aircraft motion.

$$\dot{x} = v; \quad \dot{v} = u \quad (5)$$

Here, x is the inertial position, v is the aircraft ground speed, and u is the commanded acceleration. The spacing control law shown in (3) still commands a velocity, v^c , and u is designed to track the commanded velocity using a first-order model: $u = k_v(v^c - v)$, where the gain k_v is chosen to model the response time to achieve a new speed.

Derivatives of the range error are taken to reveal the relationship between adjacent range errors.

$$e_i(t) = x_i(t - \tau) - x_{i+1}(t) \quad (6)$$

$$\dot{e}_i(t) = v_i(t - \tau) - v_{i+1}(t) \quad (7)$$

$$\begin{aligned} \ddot{e}_i(t) &= u_i(t - \tau) - u_{i+1}(t) \\ &= k_v [v_i^c(t - \tau) - v_i(t - \tau)] - k_v [v_{i+1}^c(t) - v_{i+1}(t)] \\ &= k_v [v_{i-1}(t - 2\tau) + ke_{i-1}(t - \tau) - v_i(t - \tau)] - \\ &\quad - k_v [v_i(t - \tau) + ke_i(t) - v_{i+1}(t)] \end{aligned} \quad (8)$$

Equation (8) can be rearranged using the following relationships for $\dot{e}_i(t - \tau)$ and $e_i(t)$.

$$\dot{e}_{i-1}(t - \tau) = v_{i-1}(t - 2\tau) - v_i(t - \tau) \quad (9)$$

$$\dot{e}_i(t) = v_i(t - \tau) - v_{i+1}(t) \quad (10)$$

The coupled error dynamics are shown below.

$$\ddot{e}_i(t) + k_v \dot{e}_i(t) + k_v k e_i(t) = k_v \dot{e}_{i-1}(t - \tau) + k_v k e_{i-1}(t - \tau) \quad (11)$$

The range-error transfer function is found by transforming the error dynamics to the frequency domain.

$$H_i(s) \equiv \frac{E_i(s)}{E_{i-1}(s)} = \frac{(k_v s + k_v k) e^{-\tau s}}{s^2 + k_v s + k_v k} \quad (12)$$

It can also be shown that the range-error transfer function $H_1(s)$, which relates the range error between the first two aircraft in the string to the velocity of the first aircraft, is given by the expression below.

$$H_1(s) \equiv \frac{E_1(s)}{V_1(s)} = \frac{s e^{-\tau s}}{s^2 + k_v s + k_v k} \quad (13)$$

As a comparison, consider a constant-distance control law with the range error and speed control law defined as shown.

$$e_i(t) = x_i(t) - x_{i+1}(t) - d_i; \quad v_{i+1}^c(t) = v_i(t) + k e_i(t) \quad (14)$$

Here, d_i is the constant desired distance between the i th and $(i+1)$ th aircraft. The range-error transfer function for this control law has a similar form to (12).

$$H_i(s) \equiv \frac{E_i(s)}{E_{i-1}(s)} = \frac{k_v s + k_v k}{s^2 + k_v s + k_v k} \quad (15)$$

Results in the literature show that the range-error transfer function in (15) yields a string-unstable system for all values of $k, k_v > 0$ [14].

B. Frequency-Response Magnitudes

Magnitudes of the frequency-response function are used to determine whether the closed-loop system is string stable [14]. Fig. 1 shows the magnitudes of the frequency response for the time-history range-error transfer function in (12) with different control gains; the time-delay term is modeled using a fifth-order Padé approximation. The string-unstable nature of the control law is revealed by frequency-response magnitudes greater than one. For these control forms and gains, range errors and spacing errors will propagate along the string for inputs containing frequencies in the range where magnitudes are greater than one. Although not shown here, the time-history and constant-distance control laws yield identical frequency-response magnitudes, which indicates that the time-history and constant-distance control laws yield identical string behavior. Therefore, the time-history control law is invariant with respect to the desired spacing interval and is string unstable for all $k, k_v > 0$.

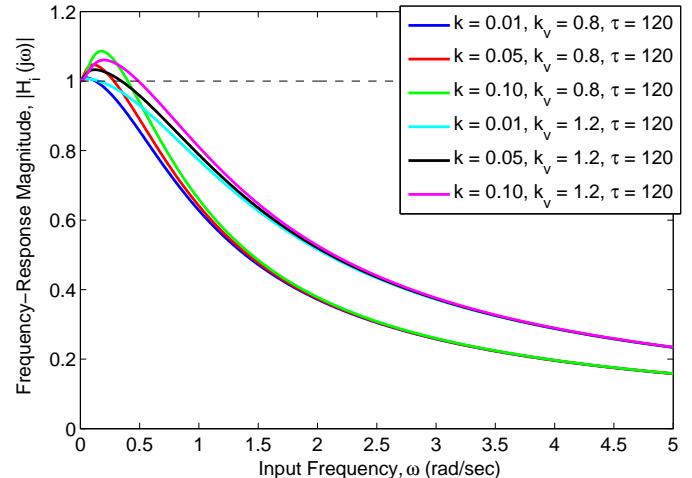


Fig. 1. Frequency-response magnitudes for the time-history control law. Different control gains are shown; all choices of gains yield string-unstable behavior.

C. Frequency-Response Magnitudes with Spacing Anticipation

An analytical expression relating adjacent range errors was not found for the time-history control law with spacing anticipation; however, the frequency-response magnitudes were numerically generated using simulations of a five aircraft string.³ Fig. 2 compares the frequency-response magnitudes for $\tau_{sa} = 0, 10$, and 20 seconds using the gains $k = 0.05$ and $k_v = 0.80$. The analytical solution from Fig. 1 is also shown to demonstrate accuracy of the numerical solution. The spacing anticipation time does not make the closed-loop system string stable, but rather changes the range of frequencies that cause string instabilities.

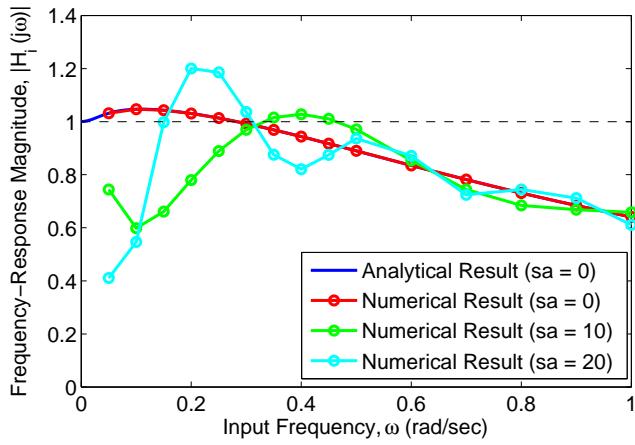


Fig. 2. Frequency-response magnitudes for the time-history control law with spacing anticipation.

IV. SIMULATION RESULTS

In this section, the assumptions for the simulation are presented including the aircraft model and approach trajectory. Simulation results are first shown for Gaussian-distributed initial spacing errors. Results are then shown for initial spacing errors that are selected not only to excite string instabilities in the system, but are also feasible initial conditions in actual operations.

A. Simulation Assumptions

1) Aircraft Model: An aircraft point-mass model is used for the simulation analysis. Because the control law is used for longitudinal aircraft spacing, the simulation is designed for straight-line, longitudinal dynamics only. The longitudinal model, simplified from a commonly used point-mass model,

³A five aircraft string was simulated using the control law in (4); the first aircraft in the string was subjected to a sinusoidal disturbance of the form $d = A \sin(\omega t)$. Magnitudes of the steady-state range errors were then used to determine the ratio of adjacent range errors for the different disturbance frequencies.

is shown below [20].

$$\dot{x} = V \cos \gamma \quad (16)$$

$$\dot{h} = V \sin \gamma \quad (17)$$

$$\dot{V} = \frac{T - D}{m} - g \sin \gamma \quad (18)$$

Here, x is the along-track position, h is the altitude, and V is the true airspeed, which is equal to the ground speed assuming no winds; m is the aircraft mass, g is the gravitational constant, and D is the drag force, which is dependent upon aircraft configuration and true airspeed. The control inputs to the aircraft are thrust, T , and flight path angle, γ . To model non-instantaneous changes in the flight path angle, a first-order model is assumed to track the commanded flight-path angle, γ_c : $\dot{\gamma} = k_\gamma(\gamma_c - \gamma)$.

Guidance control laws are designed to track the vertical profile and the commanded speed determined by the time-history control law. The Flight Management System (FMS) is assumed to be in *VNAV-path* mode, which uses pitch, or flight path angle for the point-mass model presented here, to control the aircraft to the vertical path, and thrust is used to control longitudinal speed [21].

EUROCONTROL's Base of Aircraft Data (BADA) is used to calculate the drag coefficients, stall speeds, and thrust limitations for different flight phases [22].

2) Approach Trajectory: Scenarios are simulated for an approach trajectory starting 30 nautical miles (NM) from the FAF. The initial altitude and indicated airspeed (IAS) constraints are 12,000 feet (ft) and 230 knots (kts), respectively; the final altitude and IAS constraints are 4,000 ft and 180 kts, respectively.

For a Boeing 767-300 and a range of aircraft weights, an idle descent from the initial altitude to the start of the deceleration is assumed. The deceleration segment of the trajectory assumes a 0.30 energy-share factor to decelerate to the final speed constraint. Fig. 3 shows the altitude and IAS as a function of range to the FAF for the trajectory used in the analysis.

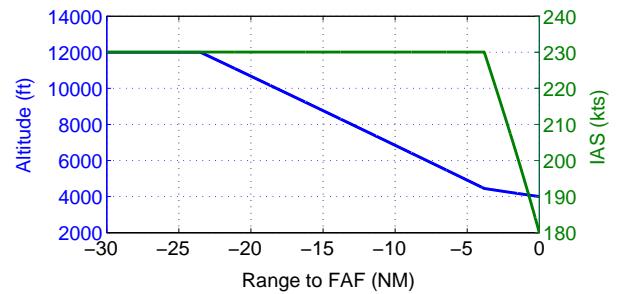


Fig. 3. Altitude (blue line) and IAS (green line) of the reference trajectory as a function of range to the FAF.

B. Gaussian-Distributed Initial Spacing Errors

A string of ten aircraft was simulated to illustrate the behavior of the nominal time-history control law in (3). The

initial spacing errors of the second through tenth aircraft are normally distributed with a variance of three seconds. It is assumed that the ADS-B messages are updated every second based upon expectations for actual operations. Therefore, the spacing aircraft has a record of the target aircraft's states each second. Linear interpolation is used to determine the target aircraft's states at more frequent intervals. Surveillance errors are not modeled in the ADS-B messages. These errors are not assumed large enough to cause an effect on overall string performance, but do have an effect on the achievable spacing at the FAF; the effect of surveillance errors is outside of the scope of this paper. The speed control law is updated every 0.1 seconds. Commanded ground speeds calculated by the control law are converted to indicated airspeeds, which are the speed commands issued to the spacing aircraft.

Fig. 4 shows the commanded indicated airspeeds and resulting ground speeds of the ten aircraft in the string as a function of range to the FAF. Commanded speeds are a mix of increases and decreases related to the initial spacing between aircraft pairs. These results assume a continuous implementation of speed commands similar to what would be expected for IM equipment that is coupled to the FMS. A continuous implementation of the speed commands leads to the spacing error performance shown in Fig. 5, where spacing errors indicate that the aircraft are converging to their desired intervals. The spacing errors, as defined in (1), are calculated every 0.5 NM along the path. The initial spacing errors are within ± 5 seconds; these spacing errors are reduced to within ± 0.1 seconds prior to the deceleration segment; and, the spacing errors at the FAF are within ± 0.15 seconds, which is well within operationally acceptable limits.

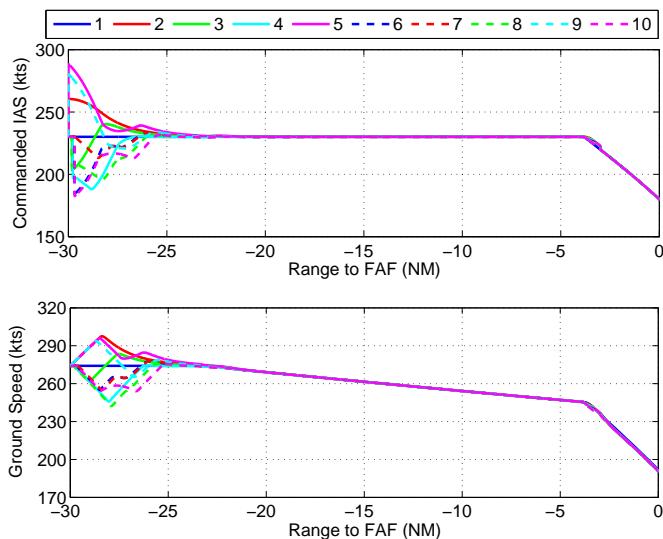


Fig. 4. Commanded indicated airspeeds and ground speeds for the time-history control law with Gaussian-distributed initial spacing errors.

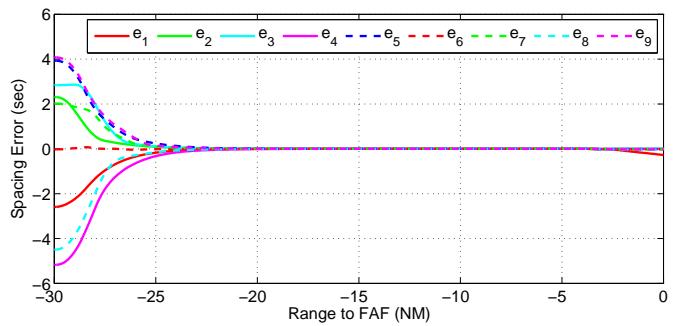


Fig. 5. Spacing errors as a function of range to FAF for Gaussian-distributed initial spacing errors.

C. String Behavior Simulations

To investigate the string behavior of the time-history control law, the second aircraft is initialized with a non-zero spacing error relative to the first aircraft, and the third through tenth aircraft have zero spacing errors with respect to their targets. These initial conditions result in a step speed command to the second aircraft, which is then passed through the string via the control law. A Fourier transform of a step command reveals that the input to the second aircraft will be dominated by lower frequencies near zero, which will excite the string instabilities shown in Fig. 1.

1) Nominal Control Law: Fig. 6 shows the commanded indicated airspeeds and ground speeds for the ten aircraft string using the nominal time-history control law in (3).⁴

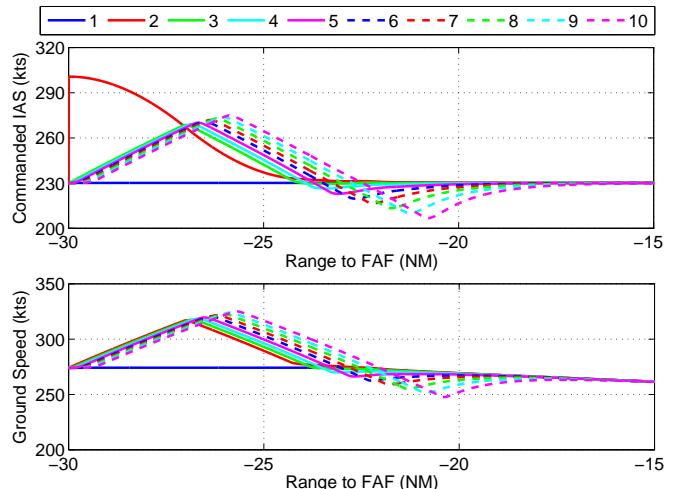


Fig. 6. Commanded indicated airspeeds and resulting ground speeds for a ten aircraft string using the nominal time-history control law.

As predicted by the closed-form analysis, string instabilities are excited due to the step command to the second aircraft,

⁴Note that Fig. 6 shows only the region of interest between -30 and -15 NM to the FAF in order to more clearly show the system response. Simulation results outside of the region shown are similar to what was observed for the case with Gaussian-distributed initial spacing errors.

which results in increasing speed commands along the string. Whereas there is only a ten knot difference in the maximum ground speeds achieved between the second and tenth aircraft, this control law would not support a long string of aircraft where speed limitations would constrain commanded speeds farther downstream of the disturbance. Following an increase in commanded speeds, aircraft are then commanded to slow down resulting in an inefficient operation, which would likely not be acceptable to flight crews, controllers, or airline operators.

The spacing errors, shown in Fig. 7, also grow along the string as evidence of the string instabilities. Despite the large speed changes, the speed control law is able to reduce the disturbance effects to precisely space aircraft at the FAF. Spacing errors at the FAF are within ± 0.2 seconds with the exception of the spacing error between the first and second aircraft, which is -0.3 seconds. This is believed to be a result of errors in tracking the commanded speed during the deceleration segment.

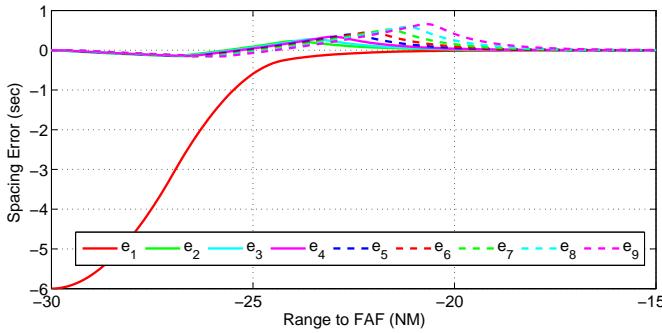


Fig. 7. Time-based spacing errors for a ten aircraft string using the time-history control law.

2) Spacing Anticipation: The effects of using a spacing anticipation time, as described by (4), are shown in Fig. 8 for $\tau_{sa} = 10$ seconds. The second aircraft again has a six-second initial spacing error relative to the first aircraft. Results in Fig. 8 show that using a spacing anticipation time in the control law yields improved string performance over no spacing anticipation. Because the step input to the second aircraft is dominated by lower frequencies, the shift in the string-unstable frequency range to higher frequencies, as shown in Fig. 2, 2, leads to the improved performance. In this case, maximum speed commands are decreasing along the string. The initial spacing error between the first and second aircraft would eventually be dissipated for a longer aircraft string and would not be evident to aircraft further back in the string.

These results are consistent with the results presented in reference [19]. However, note that the spacing anticipation time leads to undesired effects during the planned deceleration; the aircraft along the string begin to decelerate earlier and earlier in order to match the anticipated speed of their targets. This ultimately leads to poorer spacing at the FAF with spacing errors between 0 and -1 seconds, although spacing errors of

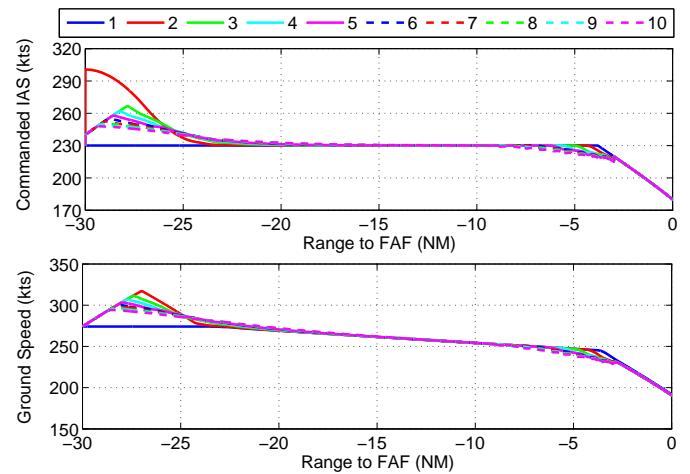


Fig. 8. Commanded indicated airspeeds and resulting ground speeds for a ten aircraft string using the time-history control law with spacing anticipation time $\tau_{sa} = 10$ seconds.

this magnitude are not operationally significant.

Fig. 9 shows the aircraft speeds for $\tau_{sa} = 20$ seconds. In this case, the correction of initial spacing errors yields good string behavior; however, the deceleration to the speed at the FAF leads to string-unstable behavior as shown by the oscillations in the commanded and ground speeds along the string. Fig. 10 shows the spacing errors for $\tau_{sa} = 20$ seconds, where the string instabilities are also evident.

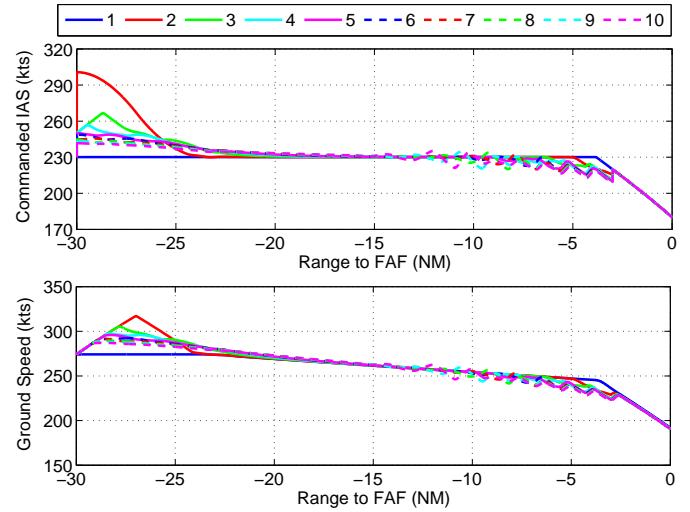


Fig. 9. Commanded indicated airspeeds and resulting ground speeds for a ten aircraft string using the time-history control law with spacing anticipation time $\tau_{sa} = 20$ seconds.

3) Spacing-Error Deadbands: Whereas using spacing anticipation is shown to improve initial spacing errors, less-efficient or even string-unstable behavior is exhibited during the planned deceleration of the target aircraft. Spacing-error deadbands are another method to improve string performance, where estimated spacing errors within a chosen threshold are

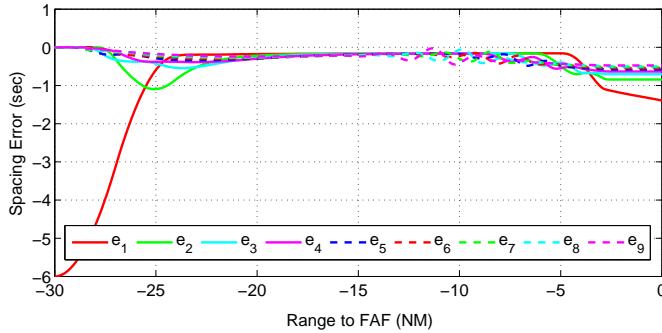


Fig. 10. Time-based spacing errors for a ten aircraft string using the time-history control law with spacing anticipation time $\tau_{sa} = 20$ seconds.

set equal to zero. Estimated spacing errors are calculated by dividing the range error by the spacing aircraft's ground speed. Fig. 11 shows the effects of using the nominal time-history control law when estimated spacing errors below a threshold of ± 3 seconds are considered equal to zero. Whereas the maximum speed commands are not drastically reduced as shown in the results for the control law with spacing anticipation, the spacing-error deadbands do prevent the growth of speed commands along the string. It should also be noted that slowdown commands are not exhibited following the increase in speed commands in response to initial spacing errors. An obvious drawback is that estimated spacing errors within the threshold are not corrected.

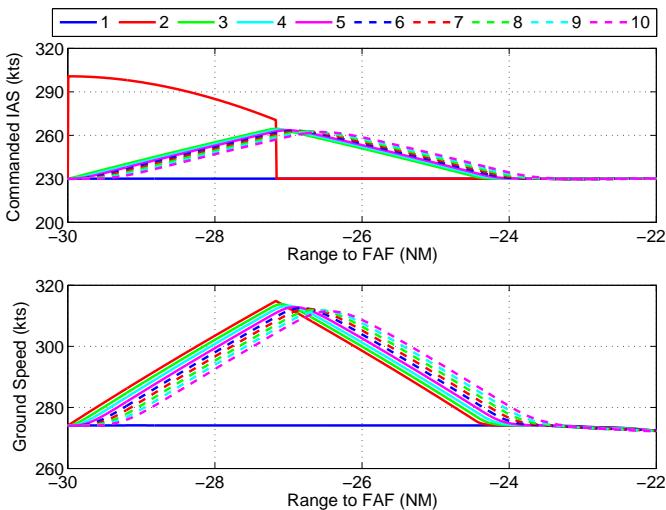


Fig. 11. Commanded indicated airspeeds and resulting ground speeds for a ten aircraft string using the nominal time-history control law with spacing error deadbands of ± 3 seconds.

4) Discrete Speed Commands: For near-term implementation, it is not expected that the IM equipment will be coupled to the FMS. Therefore, a commanded speed will be displayed to the flight crew to be manually flown. Discrete speed commands may be provided to the flight crew in five-knot

increments. Fig. 12 shows the commanded speeds and ground speeds for the nominal control law with spacing-error deadbands when the speeds are implemented in a discrete manner. It is assumed that the speeds are applied instantaneously, and future studies should evaluate the effects of flight crew delays in implementing speed commands. No significant differences in the string behavior are evident with the exception of a leveling-off of maximum speed commands along the string.

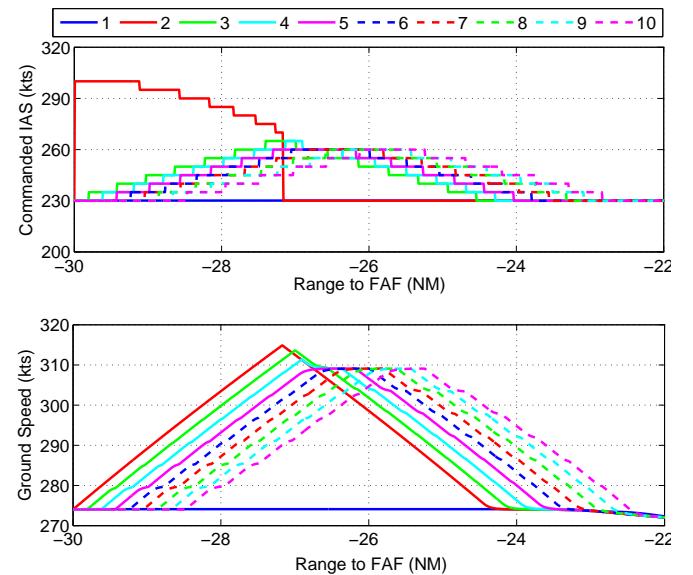


Fig. 12. Commanded indicated airspeeds and resulting ground speeds for a ten aircraft string using the nominal time-history control law with spacing-error deadbands of ± 3 seconds and discrete speed commands in five-knot increments.

D. Discussion of Results

Simulation results confirm that the time-history control law is string unstable, which results in undesired string behavior, such as increasing speed commands and spacing errors along the string. Using a spacing anticipation time to anticipate changes in the target aircraft's speed does yield improved string behavior when correcting for initial spacing errors; however, there are drawbacks in the performance when the target aircraft is decelerating. The value of the spacing anticipation time can lead to string instabilities during planned decelerations. The addition of spacing-error deadbands can improve the performance of the time-history control law without spacing anticipation; however, maximum speed commands are only slightly decreased along the string. It should be noted that these simulation results were generated for a single set of control gains, and string performance may vary significantly with variations in control gains.

Whereas it is difficult to demonstrate string stability, or good string performance, for all possible flight conditions and initial spacing errors, a practical implementation of IM operations may be to limit string lengths. By limiting string lengths,

the growth of spacing errors along the string will not result in speed saturation and inefficient or unsafe operations. In addition, the actual control law implementation may combine techniques, such as varying the spacing anticipation time from zero to a non-zero value, to yield more optimal performance for certain flight conditions. The simulation results presented here also give insight into the initialization criteria for IM operations. If a known string-unstable control law is used, a large spacing error between two strings of aircraft may cause spacing errors to be propagated through the trailing string upon initializing IM. More efficient system performance may result from keeping the two strings uncoupled. Initialization criteria defined using the type of analysis presented here can serve as a guideline for air traffic controllers tasked with setting up IM operations.

V. CONCLUSIONS

A time-history speed control law, which is a candidate speed control law for Interval Management operations, has been described and analyzed with a specific focus on string stability. A simplified, closed-form string stability analysis indicated that the time-history control law is not string stable. Numerical analysis of a time-history control law with spacing anticipation also revealed string instabilities for this control form. The value of the spacing anticipation time was shown to alter the frequency range where string instabilities occur.

Simulation results were used to evaluate string behavior for an approach-spacing operation. The time-history control law led to string instabilities for a set of initial conditions that was designed to excite string-unstable behavior. The time-history control law with spacing anticipation showed improved string performance in correcting for initial spacing errors; however, degraded performance was observed for spacing during the planned deceleration in the trajectory. The use of spacing-error deadbands reduced some of the inefficiencies in the time-history control law without spacing anticipation, but the effects of the initial spacing errors were not attenuated through the string.

Future research will evaluate a wide range of expected target aircraft speeds to characterize the resulting behavior of the spacing aircraft. That research is expected to reveal the types of trajectories and initial conditions that will excite undesired string performance. A practical result of this string stability analysis is that aircraft strings should be limited to a certain number of aircraft to prevent disturbances from affecting a large number of aircraft. Research into the optimal number of aircraft in a string to maintain desired and efficient string performance will continue.

REFERENCES

- [1] J. A. Sorensen and T. Goka, "Analysis of in-trail following dynamics of cdti-equipped aircraft," *Journal of Guidance*, vol. 6, no. 3, pp. 162–169, 1983.
- [2] A. Pritchett and L. Yankosky, "Pilot performance at new atm operations: Maintaining in-trail separation and arrival sequencing," in *Proceedings of the AIAA Guidance, Navigation, and Control Conference*, 2000.
- [3] L. Ren, J.-P. Clarke, and N. T. Ho, "Achieving low approach noise without sacrificing capacity," in *Proceedings of the 22nd Digital Avionics Systems Conference*, 2003.
- [4] J. L. D. Prins, K. Schippers, M. Mulder, M. van Paassen, A. in't Veld, and J. Clarke, "Enhanced self-spacing algorithm for three-degree decelerating approaches," in *Proceedings of the AIAA Guidance, Navigation, and Control Conference*, 2005.
- [5] W. de Gaay Fortman, M. van Paassen, M. Mulder, A. in't Veld, and J. Clarke, "Implementing time-based spacing for decelerating approaches," *Journal of Aircraft*, vol. 44, no. 1, pp. 106–118.
- [6] B. Barmore, "Airborne precision spacing: A trajectory-based approach to improve terminal area operations," in *Proceedings of the 25th Digital Avionics and Systems Conference*, Portland, Oregon, 2006.
- [7] K. Krishnamurthy, B. Barmore, F. Bussink, L. Weitz, and L. Dahlene, "Fast-time evaluations of an airborne merging and spacing concept for terminal arrival operations," in *Proceedings of the AIAA Guidance, Navigation, and Control Conference*, San Francisco, California, 2005.
- [8] T. S. Abbott, "Speed control law for precision terminal area in-trail self spacing," NASA Langley Research Center, Hampton, VA, Tech. Rep., July 2002, NASA/TM-2002-211742.
- [9] B. E. Barmore, T. S. Abbott, W. R. Capron, and B. T. Baxley, "Simulation results for airborne precision spacing along continuous descent arrivals," in *Proceedings of the AIAA Aviation Technology, Integration, and Operations Conference*, 2008.
- [10] "Cospace 2005 - ASAS sequencing and merging: Flight deck user requirements," Eurocontrol Experimental Center, Bretigny-sur-Orge, France, Tech. Rep., January 2006.
- [11] E. Hoffman, D. Ivanescu, C. Shaw, and K. Zeghal, "Analysis of constant time delay airborne spacing concepts between aircraft of mixed types in varying wind conditions," in *Proceedings of the 5th USA/Europe Air Traffic Management R&D Seminar*, Budapest, Hungary, 2003.
- [12] D. Ivanescu, C. Shaw, E. Hoffman, and K. Zeghal, "Towards performance requirements for airborne spacing - a sensitivity analysis of spacing accuracy," in *Proceedings of the AIAA Aviation Technology, Integration, and Operations Conference*, 2006.
- [13] B. E. Barmore, R. S. Bone, and W. J. Penhallegon, "Flight-deck merging and spacing operations," *Air Traffic Control Quarterly*, vol. 17, no. 1, pp. 5–37, 2009.
- [14] D. Swaroop and J. K. Hedrick, "Constant spacing strategies for platooning in automated highway systems," *Journal of Dynamic Systems, Measurement, and Control*, vol. 121, no. 3, pp. 462–470, 1999.
- [15] D. Swaroop and K. R. Rajagopal, "A review of constant time headway policy for automatic vehicle following," in *Proceedings of the 2001 IEEE Intelligent Transportation Systems Conference*, Oakland, CA, 2001.
- [16] P. Seiler, A. Pant, and K. Hedrick, "Disturbance propagation in vehicle strings," *IEEE Transactions on Automatic Control*, vol. 49, no. 10, pp. 1835–1841, 2004.
- [17] G. L. Slater, "Dynamics of self-spacing in a stream of in-trail aircraft," in *Proceedings of the AIAA Guidance, Navigation, and Control Conference*, 2002.
- [18] L. A. Weitz, "Decentralized, cooperative control for multivehicle systems: Design and stability analysis," Ph.D. dissertation, Texas A&M University, College Station, Texas, 2009.
- [19] D. Ivanescu, C. Shaw, K. Zeghal, and E. Hoffman, "Propagation of airborne spacing errors in merging traffic streams," in *Proceedings of the 7th USA/Europe Air Traffic Management R&D Seminar*, 2007.
- [20] W. Glover and J. Lygeros, "A multi-aircraft model for conflict detection and resolution algorithm evaluation," IST-2001-32460 of European Commission, Tech. Rep., September 2004.
- [21] R. Walter, *Digital Avionics Handbook: Elements, Software and Functions*, 2nd ed. CRC Press, 2007, ch. 20.
- [22] "User manual for the base of aircraft data (BADA) revision 3.7," Tech. Rep., March 2009, Eurocontrol Experimental Center Technical/Scientific Report No. 2009-003.

Towards Universal Beacon Code Assignment

Spatial and Temporal Analysis of En-route Traffic in NAS

Vivek Kumar (Ph. D Candidate)

Research Associate

Center for Air Transportation Systems Research
George Mason University
Fairfax, VA 22030, USA
Email: vkumar3@gmu.edu

Lance Sherry (Ph. D)

Director

Center for Air Transportation Systems Research
George Mason University
Fairfax, VA 22030, USA
Email: lsherry@gmu.edu

Abstract— Beacon codes are a set of very limited National Airspace System (NAS) resource. Currently, the beacon code allocation process is based upon the concept of discreet beacon code assignment to each ARTCC. Due to the mismatch between the limited number of beacon code subsets available and the volume of traffic, duplicate beacon code assignment is unavoidable under the current scheme.

In this paper, the spatial and temporal structure of en-route NAS traffic is analyzed. This analysis provides the foundation of exploiting the inherent structure of NAS traffic to enable a more efficient beacon code assignment, i.e. with fewer beacon code changes per flight.

Keywords-component; *Beacon Code, Squawk Code, ATCRBS, En-route traffic, NAS, Air Traffic Controller workload.*

I. INTRODUCTION

The catastrophic 1960 collision between United DC-8 and TWA super constellation called for major technological overhauling of the existing Air Traffic Management system. Even though the collision was attributed to the fault of pilots, it was also partially blamed on the lack of secondary surveillance radar, which, if present, could have helped alert the controllers detect impending collision and alert the pilots. In 1961, President Kennedy ordered FAA to begin to “conduct a scientific, engineering overview of our aviation facilities and related research and development”. As a result of this order, “Project Beacon” committee was formed to investigate the deficiencies in the air traffic control system and propose alternate solutions. This resulted in the development and installation of ATCRBS (Air Traffic Control Radar Beacon System) as a new air traffic control technology.

ATCRBS is an acronym for Air Traffic Control Radar Beacon System. It is a system used in ATC to enhance surveillance radar monitoring and separation of aircraft[1][2]. ATCRBS consists of transponders (in aircraft) and Secondary Surveillance Radar (SSR) which is co-located with the Primary Surveillance Radar (PSR) on the ground. The SSR located at the ATC site, transmits interrogations and listens for replies. Transponders located on the aircraft receive interrogations, decode it and respond with requested information (mode 1,2,3/A,C). Note: Mode 4 and mode S are not part of the ATCRBS system even though they use the same transmit and receive hardware. ATCRBS interrogator at the ATC facility on

ground periodically interrogates aircraft on a frequency of 1030 MHz. Aircraft receiving this interrogation reply with the requested information (altitude and/or identification) after a 3 micro second delay. The interrogator then decodes the reply and identifies the aircraft.



Figure 1. Representation of 500 flight tracks in CONUS on 18th July 2007 in Google Earth.

When an aircraft receives a mode 3/A interrogation, the reply expected is a Beacon/Squawk code. Current mode 3/A transponders installed on aircraft are designed to transmit four octal digits, resulting in a total of $84 = 4096$ possible beacon codes. Many of the codes are reserved for special uses such as military operations, which further reduces the number of codes available for use by civilian aviation. The National Beacon Code Allocation Plan (NBCAP) established by DoT/FAA order 7110.66C [3] permanently allocates the remaining beacon codes to the ARTCC's. The controller uses the beacon code as the unique identifier for flights within a center boundary. In order for the controller to uniquely identify and address each aircraft it is necessary to ensure that all aircraft flying within the area of responsibility of that controller are uniquely identifiable, i.e. each aircraft within that center has a unique beacon code.

A. Beacon Code Allotment process

A flight is assigned its first beacon code by the Host Computer System (HCS) of the departure center. Ideally, flights could fly from their origin to their destination using the

same code for the duration of their flights. However, because there are more flights in need of codes at any point in time than there are codes available, and traffic levels are growing, each code has to be assigned to more than one flight (while still ensuring unique code assignment within a center). Due to this shortage of available codes, when aircraft cross center boundaries along its route, there is a strong possibility that the code it is using is already in use by some other flight in the facility it is approaching. The HCS in this case must reassign the flight entering the center a new beacon code. This process is known as code reassignment. This process requires voice communication to request the new transponder setting. As a result, these actions lead to increased workload of air traffic controllers and pilots (in addition to other tasks to be completed as part of a standard hand-off procedure). Therefore, code reassessments are undesirable and should be avoided as much as possible.

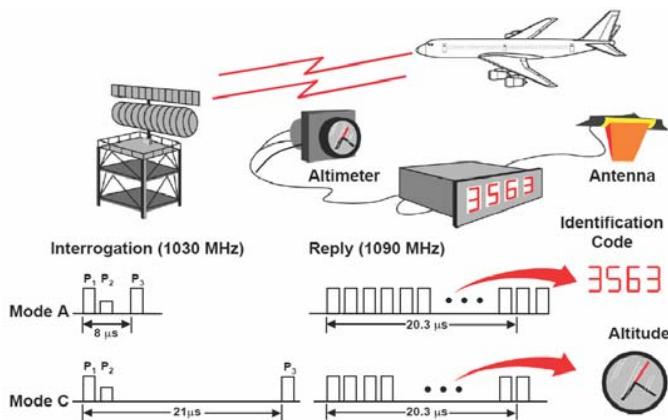


Figure 2. ATCRBS system working, Ref: Surveillance Tools for Distributed Air/Ground Traffic Management (Steven R. Bussolari, MIT, Lincoln Labs)

B. NBCAP (National Beacon Code Allocation Plan) – Primary Goal and Concept

The primary goal is to efficiently manage the beacon code set as a limited National Airspace System (NAS) resource [3]. The NBCAP is based upon the concept of discrete beacon code assignments to each ARTCC so that codes can be adapted and assigned by a computer to a flight plan according to a specific procedure. Ideally, each ARTCC should be allocated enough exclusive code blocks so that each aircraft could be given a computer assigned unique discrete code which would not be duplicated anywhere in the NAS. The intent would thereby, allow all aircraft to proceed from departure to destination using the same discrete code. Unfortunately, duplicate computer code assignments are unavoidable because of the limited number of code subsets available, the number of ARTCC's, and the volume of traffic. To minimize the impact of duplicate computer assignments, careful analysis of code utilization statistics is required to ensure appropriate facility assignments. Therefore, ARTCC facility assignments are managed from the national level.

The objective of this research is to analyze historical data to extract spatial and temporal pattern(s). This analysis would help in understanding the beacon code assignment problem

complexity at different levels (entire NAS, small geographic or time concentration).

This paper is organized as follows: Section II describes the previous work in this field. Section III describes the source of data and the relevant fields. Section IV describes the method of analysis and the results. Section V discusses the implications of these results.

II. PREVIOUS WORK

The most prominent previous work on beacon code allocation was done by Lucic et al at CSSI [4]. This research focused on the optimization of beacon code allocations based on a new geographic scheme. When assigning a code to a flight, the current beacon code assignment procedure does not consider the flight parameters or the decisions of the other Host Computer Systems in the NAS. Consequently, a code assigned to a flight in one center may end up being used in other centers in the NAS. The current (baseline) code allocation scheme works well when each center has a unique set of allocated codes. However, the limited number of initially available beacon codes combined with growing traffic requires each code to be allocated to more and more centers. Beacon codes allocated to multiple centers and high traffic levels increase the number of code reassessments and consequently further increase demand for code allocations. To solve the problem, a geographic beacon code allocation scheme was proposed. It considers the destination region of a flight while the beacon code to be assigned to the flight is being selected. This further allows the allocation of beacon codes to centers in a way to be used for assignment to flights heading toward specific destination regions. The code allocation was developed based on multiple days of ETMS data. The data were initially used to estimate code demand and to determine the interference between center-regions. Since the code allocation to center-regions consists of primary and secondary blocks of codes, two optimization problems were defined. The primary code allocation is a set of codes to be assigned to the traffic with the highest priority, it was determined for all center-regions first. The center-regions' primary code allocation optimization is aimed to allocate the available codes proportionally to center-regions' code demands while allowing small or no interference between center-regions sharing the code allocation. Since each center-region needs a specific number of codes to support the traffic, the difference between the required number of codes and size of primary allocation is allocated in the secondary block of codes in a way that minimizes code sharing between center-regions with high interference.

The proposed allocation was tested using the beacon code assignment simulation. A total of 31 days of ETMS data were included in the simulation testing. Test samples were taken from different years and months to capture traffic variations. The test results show that the proposed allocation reduces the total number of reassessments by approximately 60% with standard deviation of approximately 2%.

III. DATA SOURCES

The data used in this study was obtained from ETMS (Enhanced Traffic Management System). The data extracted

spans over a period of 24 hours starting from 4 AM EDT on 18th July, 2007 to 4 AM EDT on 19th July 2007. A snapshot of the data is shown in Figure 3. below.

FID	Time	ACID	AcType	Ori	Dest	Lat	Lon	Alt
547236635	28893	AAL256	B752	ORD	EINN	2906	3674	370
547236635	28953	AAL256	B752	ORD	EINN	2909	3662	370
547236635	29013	AAL256	B752	ORD	EINN	2912	3648	370
547236635	29073	AAL256	B752	ORD	EINN	2915	3634	370
547236635	29133	AAL256	B752	ORD	EINN	2918	3621	370
547236635	29193	AAL256	B752	ORD	EINN	2921	3609	370
547238499	30582	AAL900	B772	SAEZ	MIA	1427	4703	400
547238499	30642	AAL900	B772	SAEZ	MIA	1435	4706	400
547238499	31422	AAL900	B772	SAEZ	MIA	1515	4776	179
547238499	31482	AAL900	B772	SAEZ	MIA	1520	4781	158
547238499	31542	AAL900	B772	SAEZ	MIA	1525	4786	138
547238823	29100	COA30	B762	SBGR	EWR	1550	3936	360
547238823	29700	COA30	B762	SBGR	EWR	1626	3971	360
547238823	29820	COA30	B762	SBGR	EWR	1634	3974	360
547238823	30360	COA30	B762	SBGR	EWR	1699	4005	360

Figure 3. Snapshot of enroute ETMS data for 18th July 2007

The data has 48966 tracks (each track is one flight leg). It also includes international and cargo flights. Military flights are not included in this data as they have their own reserved set of beacon codes. The following fields were used:

- FID – this field is the unique identifier for a flight leg.
- Time – Time in seconds from 12 AM GMT on 17th July, 2007.
- ACID – Airline assigned aircraft ID. Eg: AAL900
- AcType – Aircraft Type. Eg: B752
- Ori and Dest – 3 or 4 letter ICAO code for origin and destination airport respectively. Eg: ORD
- Lat and Lon – Latitude and Longitude of the aircraft at the corresponding time in minutes. Eg: 2906 represents 2906/60 ~ 48.43 degree North
- Alt – Represents the flight altitude level. Eg: 370 translates to 370000 feet above MSL (mean sea level).

As this analysis is focused on CONUS (Continental US), all points outside an artificially constructed imaginary boundary (15 N 50 N && 60 W 128 W) were excluded. This boundary is shown in Figure 1. in red.

IV. ANALYSIS AND RESULTS

A. National flight count

Figure 4. shows the count of active flights in the entire NAS for every 15 min time interval starting at 4 AM EDT on 18th July 2007. A flight is counted as active for all time periods between its actual gate push back (request beacon code) and gate in time (release of beacon code). If beacon codes are assigned centrally for the entire NAS, then this count of active flights on a national level for every 15 min time period would help ascertain the level of shortage of beacon codes, i.e. when do we run out of beacon codes for first time during the day?

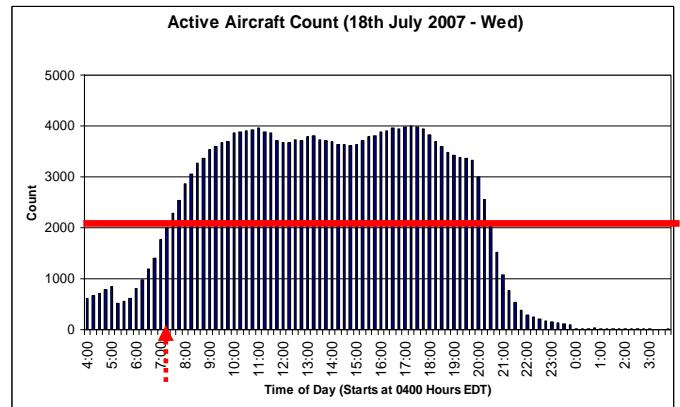


Figure 4. Active aircraft Count for Entire NAS

The active count of aircraft exceeds 2000 for the first time during the 7:00 – 7:15 AM (EDT) quarter of the day. This means that if we were to do a FCFS (First Come First Served) allotment of Beacon codes on a national level, then we would be out of codes by 7:00 AM EST. Clearly, this is not a feasible option.

B. Flight Count by geographical division

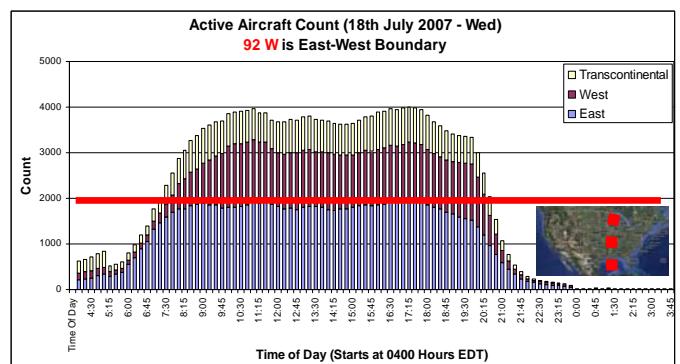
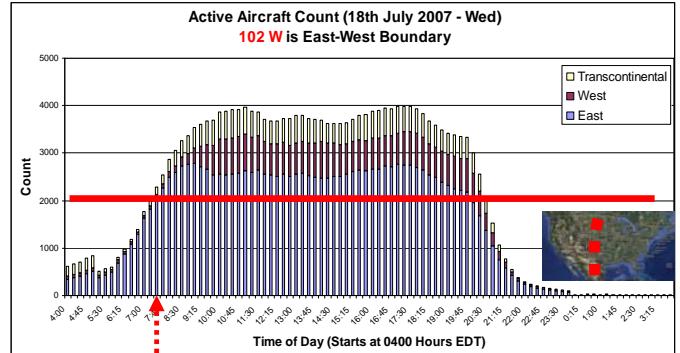


Figure 5. Active aircraft Count divided by geographical location of flight trajectory. Figure 5.a – Dividing Longitude is 102 W. Figure 5.b – Dividing Longitude is 92 W

If the entire NAS is divided into 2 halves, namely Eastern and Western corridors, then these corridors could simultaneously use the same set of codes without causing any conflicts as the flights would be geographically wide apart. For simplicity, there are two divisions of longitude considered, namely 102° W and 92° W. 102° W is Kansas/Colorado

border. Figure 5. A and B shows the count of Eastern, Western and transcontinental flights for these longitudinal divisions. With the East-West dividing boundary set as 102° W, the active aircraft count for Eastern corridor exceeds 2000 for the first time during the 08:30-08:45 AM (EDT) quarter of the day. And for the Western corridor, the active aircraft count never exceeds 769. The implication of this result is that if 102° W is used as the dividing boundary, then 2000 codes would not be sufficient for the entire Eastern corridor by itself.

However, with the boundary set as 92° W, the active count in both the eastern and western divisions remain under 2000 at all times.

C. Flight count by distance flown

Flights which travel smaller distances would relinquish their codes faster. This implies that if sufficient number of codes were reserved for short-distance flights then they could automatically be de-conflicted with the long-distance flights. To test this hypothesis, we first plot the histogram for the flight count by flight distance travelled.

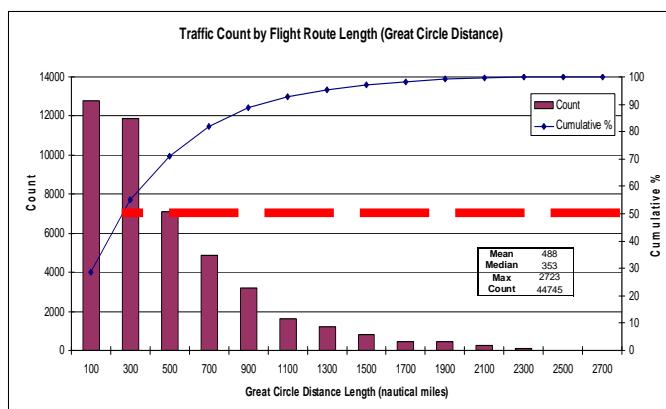


Figure 6. Distribution of Aircraft count by flight distance

Figure 8. Figure 6. shows the count of aircraft by distance travelled in miles. It can be seen that 50% of the flights travel 400 nautical miles or less. Figure 7. shows the distribution of these short and long-haul flights throughout the day. The activity count of short-haul flights (<400 nautical miles) on the East and West corridors never exceeds 1480 and 483 respectively. Similarly, it is also observed that the activity count of long-haul flights (>400 nautical miles) on the East and West corridors never exceeds 1196 and 279 respectively.

This implies that if a set of beacon codes were reserved for short-haul flights on the Eastern corridor, then at least 1480 codes would be required to make sure there is no shortage of codes at the highest traffic level period. Also, at least 1196 codes would be required for the long-haul flights on the Eastern corridor. Together this adds up to 2676 codes, which is well over the 2000 available quota of beacon codes. Therefore distance based beacon code assignment is not a viable alternative in itself. However, it might be used in combination with other heuristic procedures. The feasibility of such assignment is yet to be determined and is part of the future research.

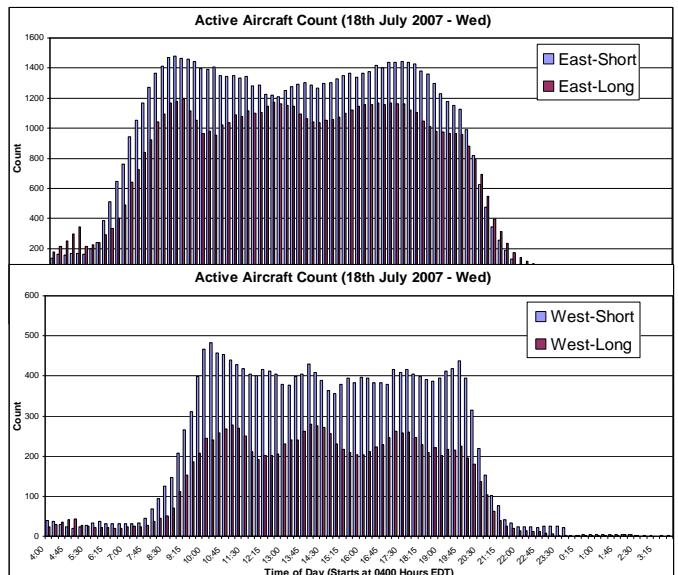


Figure 7. Distribution of Aircraft count by flight distance and geographical location (Short is defines as flight route < 400 nautical mile)

D. Sector traffic Intensity

Figure 8. shows the cumulative count of aircraft for each ARTCC for the entire 24 hour period. This data set has 11904 tracks (each track is one flight leg). ZDC with 8395 flights during the day is the ARTCC with highest cumulative traffic count, followed by ZTL and ZOB.

Figure 9. shows the traffic count for every 15 min time bin for the top 7 ARTCC. Interestingly, even though ZDC has the highest cumulative traffic count for the entire day, it has the highest instantaneous traffic count (15 min bin) for only 22 of the 96 time periods. ZTL has the highest count for 42 time periods. ZAU has the third highest number of peaks with 12 time periods corresponding to highest instantaneous traffic count.

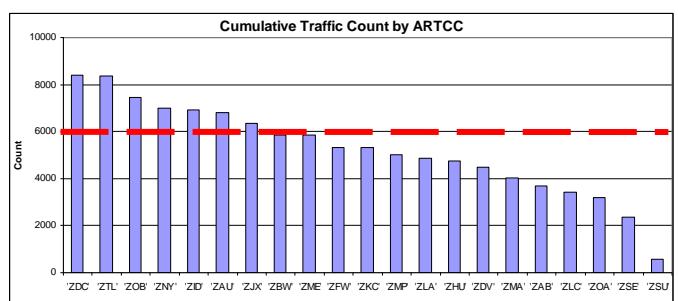


Figure 8. Cumulative count of aircraft for each ARTCC for the entire day

This implies that if mutually exclusive sets of beacon codes were assigned to ARTCC's based on the dynamic traffic intensity by time of day, then ZTL for example would need to be assigned at least 537 codes for 23rd quarter of the day. It must be noted here that more data needs to be analyzed to take into account the effect of seasonality and traffic variations due to weather for example.

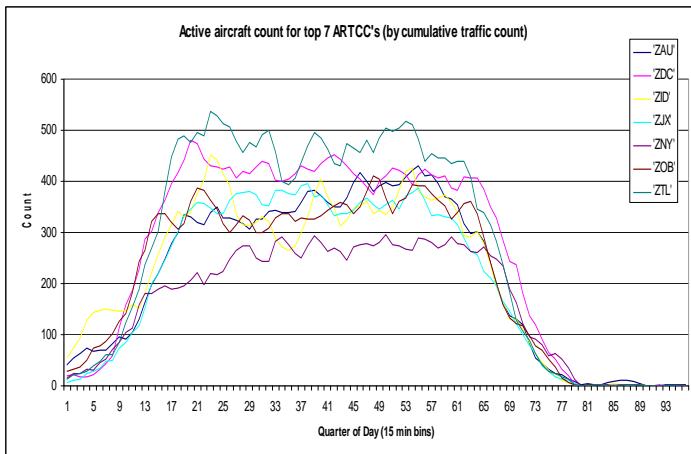


Figure 9. Cumulative count of aircraft for each ARTCC for the entire day

V. CONCLUSIONS AND FUTURE WORK

This paper presented a spatial and temporal analysis of a day's worth of 4-D trajectory data for the entire CONUS. At the national level, the maximum active flight count during the day is ~4000. As the number of available beacon codes is about 2000, this statistic precludes the possibility of national beacon code assignment without any space and time considerations.

When the CONUS is divided into two halves along an imaginary longitudinal boundary (92 W), the maximum count of active flights in the east-only and west-only category is about 1980 and 1400 respectively. As these two regions are geographically independent, they could share the same beacon code sets without any conflicts. However, flights which cross this imaginary longitudinal boundary are not included in this count and they would conflict with both the eastern and western halves.

When the flights are categorized based on distance flown and/or their geographical locations, the counts are as follows. At its peak traffic period, there are 1480 and 1196 short (<400 nautical miles) and long-haul flights on the Eastern corridor.

This adds up to 2676, which is well over the 2000 available quota of beacon codes. This leads to the conclusion that distance based beacon code assignment is not a viable alternative in itself.

The beacon code assignment problem can be formulated as mixed binary integer optimization. One of the possible objectives of such a problem could be reducing the number of code reassessments for flights. Constraints would include: Ensuring that every flight gets assigned exactly one beacon code for each time period during which it is active. Ensure that no two 'space-time adjacent' flights (flights present in the same sector at the same time period) share the same code.

ACKNOWLEDGMENT

The authors would like to acknowledge the contributions and help from the following persons and institutions.

This research is funded by FAA. The authors would like to thank Richard Jehlen for his suggestions and comments. From George Mason University the authors would like to acknowledge the help from Dr. George Donohue and Dr. John Shortle. From Metro Aviation, the authors acknowledge the help and contribution of Jason Burke and Dr. Terry Thompson.

REFERENCES

- [1] R. De Neufville, *Airport system: planning design, and management*, New York: McGraw-Hill, 2003.
- [2] E. Williams, "Airborne collision avoidance system," *Proceedings of the 9th Australian workshop on Safety critical systems and software - Volume 47*, Brisbane, Australia: Australian Computer Society, Inc., 2004, pp. 97-110.
- [3] DOT/FAA Order 7110.66C, National Beacon Code Allocation Plan NBCAP, Federal Aviation Administration
- [4] Lucic et al, "Improving Beacon Code Assignment Process", CSSI internal report

Track 8

Human Factors

The Structure and Color Optimization Process to Generate Metro-like Maps of Flight Routes

Christophe HURTER

DSNA/DTI R&D
IRIT7, avenue Edouard. Belin
31055 Toulouse France
christophe.hurter@aviation-civile.gouv.fr

Mathieu SERRURIER

DSNA/DTI R&D
IRIT7, avenue Edouard. Belin
31055 Toulouse France
serrurier@irit.fr

Roland ALONSO

DSNA/DTI R&D
IRIT7, avenue Edouard. Belin
31055 Toulouse France
alonso@cena.fr

Abstract— Aircraft must follow strict Air Traffic Control (ATC) rules. One of these rules is that aircraft have to fly over pre-defined Flight Routes (FR). Current ATC visualizations do not display FRs because they are numerous and run into each other, and thus spoil the visualization. The schematic views for metro maps are used to maximize the transmission of relevant information (lines, metro stops) of network visualization. In this paper, we will focus on two different issues. First, we show how we transposed mathematical constraints used to produce metro maps into the specific field of ATC. The view produced is a context compatible, 2D picture of a metro-like view for Air Traffic Control. Second, we propose to investigate the generation and placement of colors to be assigned to lines of the network. The first step is to find as many colors as lines of the network. These colors must be perceptually as distinct as possible, and available in the vocabulary of colors. The second step is to solve the NP-complete problem of the optimal assignment of these colors so that close lines have the most perceptively distant color. Finally, we assess the map produced through experimentation to validate its quality.

Keywords-component; *Visualization, metro map, colors assignment, Air Traffic Controller.*

I. INTRODUCTION

Fundamental research in visualization is concerned with the impact of presentation on visual perception and understanding [22] [23]. Current Air Traffic Control (ATC) environments employ complex visualization systems. These visualizations display large quantities of information that must be understandable with the minimum cognitive workload. As traffic increases together with safety levels, The ATC environment requires new kinds of visualizations.

The activity of Air Traffic Controllers (ATCos) consists in maintaining safe distances between aircraft by giving clearance to pilots (heading, speed, or altitude orders). The traffic is planned in advance: companies must request a flight plan from the regulatory authorities, which is translated into a mandatory Flight Route (FR) *i.e.* a sequence of beacons (named locations on the ground), with an associated time and altitude. A first task associated with the FRs is the a posteriori analysis of the traffic in order to determine the most conflicting FRs. Moreover, future ATC systems promote the use of FR as their main component. For instance, ATC regulators envisage a system where airline companies have to book time-slots along

FRs. A natural visualization for this kind of task is the geographical view. However FRs are numerous and run into each other. A direct representation of the FRs is too complex and unusable. An efficient visualization of FRs would simplify their utilization.

In this context, metro map visualization appears to be a suitable solution. Metro maps are schematic drawings of the underlying geographical network that represents the different stations and lines of a metro system [6]. Automatic generation of metro maps is an active, ongoing research area [16] [20]. Most existing methods consist in optimizing the shape of the networks by considering the mathematical criteria the authors want to minimize (network density, straight lines...).

Our approach is divided into two parts: The geographical optimization and the color assignment of FRs.

In the first part of this paper, we study how to adapt existing metro map generation to the ATC context. Most existing metro maps are produced manually: while metro-map design is time-consuming, it is also limited by the number of metro lines, and by the fact that the metro network does not often change. On the contrary, a metro-like view of FRs must be generated automatically, as FRs are too numerous (more than 600 FRs in the French airspace) and change regularly (their life-time can be as low as 40 days, and adding a new flight route is cheaper and physically less challenging than creating a new metro line). In order to adapt the metro map generation, we compare differences between metro and FRs layouts. Then we define specific mathematical cost functions that measure the quality of a metro-like view of the FRs. These functions are finally used in an optimization algorithm to generate optimal configurations.

The second part of our paper concerns the color assignment of the lines of a metro map. The schematic views for metro maps are used to maximize the transmission of relevant information (lines, metro stops) of network visualizations. Automatic generation of metro maps focuses principally on the physical structure of the network, but less on the color choice which is, nevertheless, an accurate visual discrimination variable. The FR discrimination will increase if close FRs have distant perceptual colors. A random color assignment is not optimal for two reasons: all the colors do not have the same differentiation power and distant FRs are already visually separated and do not need the color differentiation. Hence, we

use an algorithm of color generation and color assignment. Our goal is to produce a set of N discriminating colors to assign for N FRs. In order to facilitate the communication between users, we favor colors that can be named in the common color language. The color assignment is an NP complete problem; therefore we use simulated annealing to find an optimal solution [9].

The remainder of this paper is organized as follows. We first recall the state of the art of metro-like automatic generation. We present a description of a specific activity (Air Traffic Controllers with Flight Route management), and highlight differences with public transportation. We show how we adapted the algorithm of automatic metro-map generation by introducing new mathematical cost functions in an optimization process. Then we explain how we create a color set and how we assign it to FRs. Finally, we perform experimentation to assess the views produced.

II. ATC/METRO MAP COMPARISON

Metro maps are schematic drawings of the underlying geographical network that represents the different stations and metro lines of a metro system [6]. Ever since the first one was created in 1934 by Beck, the drawing rules have not changed: transport lines are straightened and restricted to horizontals, verticals and diagonals are depicted at 45 degrees (coined as *octilinear* in [16]). The scale in crowded downtown areas is larger than in the less dense suburbs in order to create more uniform distances between adjacent stations. In spite of all this distortion, the network topology and the relative position between metro stations must be retained. Metro maps have been tentatively applied to other classes of problems than metro traveling, such as understanding the structure of a thesis, or the sequence of teaching courses [15].

Automatic generation of metro maps is an active, ongoing research area, with no complete and general solution. The most frequently used methods are to optimize the shape of the networks by considering the mathematical criteria the authors want to maximize [16] [20]. The best results are performed by using meta-heuristic techniques such as simulated annealing. The criteria used are essentially based on the direction and bending of the line in order to generate metro maps where lines are as straight as possible and follow octilinear directions. Some methods are focused on specific criteria such as line crossing minimization [2], or path simplification [12]. A related problem is the automatic generation of directions [1]. Though not strictly a metro map, the authors have designed an activity-driven visualization, with multiple kinds of graphics and textual information.

ATCos manage air traffic in an area called a sector, (white area in Figure 1). Airways are the straight segments that planes have to follow (solid and dashed blue lines in Figure 1). Two beacons delimit an airway. A Flight Route is a sequence of airways (example: the route displayed as a thick red line is made up of three segments, and passes over 2 beacons). FRs span multiple sectors, but ATCos only need to know the section of FRs that cross their sector. Each flight has an associated FR. FR are often re-used by companies, either for regular flights (5 per day between Paris and Toulouse for

example), or by companies that compete on the same route. With the increase of the traffic some new tasks arise. The first one is the *a posteriori* analysis of traffic in order to determine the most conflicting Flight Routes and the densest sections (e.g. Flight per hour). The second is a future task for companies. It has been envisaged to ask the companies to book time slots along flight routes in order to optimize the traffic. For these two tasks a geographical representation of the sectors is the most appropriate solution. However, direct representation of FRs is not suitable due to their number and to their density in some areas. The exact location of beacons is not compulsory since this task does not require real-time traffic management. Users need an easy-to-read view of flight routes (with a low cognitive workload). In this way, the view produced can be used to understand the structure of the FRs, and to display the traffic density (e.g. by adding a size proportional to the traffic of each section of the FR).

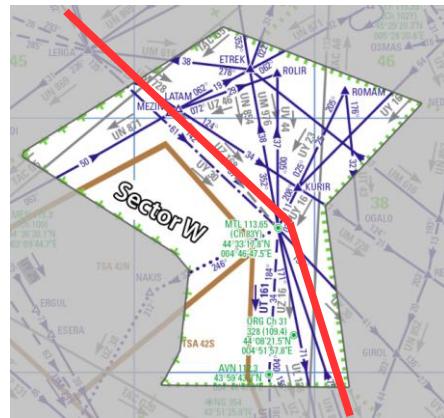


Figure 1 : a sector (white zone), and a route (red line)

By comparing problems and questions in public transportation and ATC, we can see how related they are. A common theory in the field of cognitive psychology is that people think of a path as a sequence of steps [1] [5]. For ATCo, these steps are the beacons [11]. We think that metro map-like visualization is a promising visualization for ATC. However, the metro map drawing rules used for public transportation cannot directly be applied to the ATC field. The specificity of ATC lies in the following aspects:

Subway stations vs. beacons: stations are spread out to maximize the servicing of an area by public transportation, whereas beacons are very close together. Furthermore, beacon layout is much more heterogeneous than station layout.

The regularity of station spacing: for effective service purposes, metro stations are inherently spaced regularly along a line. Thus, the constraint that equalizes distances between stations modifies only slightly the topology of the network. On the contrary beacons are not spaced regularly along the FR. A constraint that equalizes distances between beacons is not desirable, as it would alter the topology of the network to an excessive extent.

There are more “interchange” beacons than “intermediate” beacons, as a lot of FRs intersect at the same beacon. Metro stations seldom connect more than five lines.

Flight Routes have a lot of crossings compared to subway lines. This is due to the general orientation of airways (East-West and South-North flows).

Flight Routes are numerous and they often share the same sections. When FRs share the same altitude, it is as if metro lines shared the same rails.

These differences lead us to define new specific mathematical rules (cost functions) to produce a metro-like view of Flight Routes.

III. GRAPH OPTIMIZATION

Currently, the visualization of a sector and its FRs is produced manually. There is no significant previous work in the ATC field that tries to automatically produce visualizations of schematic FRs.

The graph of the representation of a sector is constructed from a set of beacons and FRs. A beacon has a name, a center position and a span value that describes the distance between each FR that connects with it. We also need the input and output location of the FR at each given beacon. The W sector (Figure 1) is composed of 19 beacons and 24 FRs. **Figure 5** and **Figure 7** are representations of FRs before the geographical optimization. As we can see, numerous beacons overlap, and FRs cross so much that it is difficult to follow a specific FR.

Each of the above design constraints can be transposed into mathematical cost functions. Cost functions can be integrated into an optimization algorithm to obtain an optimal metro-like view with respect to these visual constraints. For ATC metro maps, we consider three major constraints:

1/ *Topological closeness*: it indicates the relative closeness of represented beacons to their original position,

2/ *Density*: it indicates the number of neighboring beacons,

3/ *Crossing*: it indicates the number of crossing FRs.

Topological closeness is inherited from the classical metro maps. *Density* and *crossing* are more specific to the ATC field. *Linearity* (straight line) and *octolinearity* (rounded up angle values) are not very pertinent in the ATC context since the direction of a FR changes at each beacon.

A. Topological closeness: $Cost_{dist}$

The visual location of each beacon has to be as close as possible to the real beacon location. If the view produced significantly alters the topology, this may create a gap between the mental and physical representation of the ATC sector. In metro maps, metro stations are often inherently evenly spaced, and topological closeness is usually respected even if not constrained. Here, $Cost_{dist}$ exponentially increases with the distance between the real beacon location and its representation. $Cost_{dist}$ is defined as follows:

$$cost_{dist} = -\frac{\sum_{b \in B} erf(\frac{d(b,b') - MAX_{dist}}{C_{dist}})}{|B|}$$

Where B is the set of beacons, b and b' are respectively the initial and the modified position of the beacon. $D(b,b')$ is the Euclidean distance between b and b' . $|B|$ is the cardinal of B . Erf is the Gauss error function (see figure 4). C_{dist} and MAX_{dist} are constants respectively fixed at 200 and 30. Roughly speaking C_{dist} corresponds to the displacement that is accepted without a penalty and MAX_{dist} is the ratio of the penalty increase when the accepted displacement is reached.

B. Density: $Cost_{den}$

As pointed out in the introduction, one major ATC specificity is the high density of beacons in some areas. This density is incompatible with a clear representation. With a given beacon, $Cost_{den}$ increases with the proximity of its neighbors.

$$cost_{den} = \frac{\sum_{b_i, b_j \in B, i \neq j} erf(\frac{d(b_i, b_j) - MAX_{den}}{C_{den}})}{|B|^2}$$

Where C_{den} and MAX_{den} are constants respectively fixed to 80 and 6. Roughly speaking MAX_{den} is the minimum distance needed between the beacons. $b_i, b_j \in B, i \neq j$ is the set of all the different beacon couples.

C. Crossing: $Cost_{cross}$

Once again, a clear view requires a minimum of overlapping FRs. $Cost_{cross}$ is the rate of crossing segments. This issue has already been addressed in classical metro-map drawing, but in our context this criteria need a stronger optimization.

$$cost_{cross} = -\frac{\#crossingedges}{\#edges^2}$$

IV. ALGORITHM

According to the above considerations, an efficient view must minimize all the previously listed costs. This cost minimization is a complex problem since all these costs are interdependent. Thus this optimization problem is intractable using a deterministic algorithm. Nowadays, the best automatic metro maps are produced with meta-heuristic methods. In this paper, we use a method that is deeply rooted in optimization technique. Simulated Annealing (SA) is a powerful, general, optimization technique [9]. SA is computationally costly, but this remains acceptable since the time span to produce a metro-like view is considerably shorter than the lifetime of the visualized FR.

Cooling parameters of Simulated Annealing is tuned to 0.9997. The algorithm stops when it converges, i.e. 100 steps without increasing the function. The optimization of the sector W took approximately 3 hours on a Mac pro under Linux.

We minimize the following costs:

$$Cost_{phase1} = C^*(Cost_{den} + Cost_{dist}) + Cost_{cross}$$

Where C is a constant fixed empirically at 5. This minimization is performed by the SA. The SA is based on the notion of neighbor which corresponds, roughly speaking, to a small variation of the current state. In our case, a neighbor of a current visualization is obtained by moving the location of a beacon, changing the span of a beacon or switching the FR order of a beacon. The visualization obtained is the best trade-off between $\text{Cost}_{\text{dist}}$ and Cost_{den} which are antagonist. Parameters used in $\text{Cost}_{\text{dist}}$ and Cost_{den} create an area where these costs do not change (e.g. when beacons are not close and not too far from their initial location). In this area, the configuration with the smaller number of crossings is computed.

The initial view of the W sector is presented in **Figure 5** and in **Figure 7**. The geographically optimized view of the W sector is presented in **Figure 6** and in **Figure 8**. We can see that the view is well spaced out. The number of crossings has been lowered. The remaining crossing segments would require too much modification of the topology. Having optimized the geometry of the metro maps, we will now choose and assign the colors associated to each FR.

V. PLACEMENT AND MAP GENERATION ASSESSMENT

Most metro maps are created by a graphic designer who chooses the appropriate design and color of each metro line. The designer creates a harmonious color set with graphic techniques [7] [18] [19] [24].

In this section we propose a generic method to create color sets based on designers' techniques and previous works [8] [10] [13]. Colors must be as distinguishable as possible. The generation of distinguishable colors in the RGB space is not suitable since Euclidean distance in this space cannot be considered as a perceptual distance. Indeed, color differences in RGB space are not homogenous with the color difference perception of human beings. The only way to set up colors and assess their perceptive distances is to use the findings of the CIE (International Commission of Illumination) [4]. The CIE built, with empirical methods, models that describe the chromatic colors, taking into account human visual perception. The CIE $L^*a^*b^*$ model (CIELAB) is a color space that is homogeneous with human perception. The Euclidean distance between two colors, in the $L^*a^*b^*$ space, can be computed with de Deltae 2000 formula. The LCHab color space is the result of the transformation of Lab space in a cylindrical coordinate system. This color space uses three intuitive dimensions: Luminosity, Chroma (i.e., saturation) and Hue. L, C, H dimensions are orthogonal and perceptually uniform. However, the use of the CIELAB model is limited to the RGB gamut (actual displayable colors) of the output device.

With the CIE LCHab, we can easily compute any color set regularly spaced on the circular Hue axis. The division of the hue axis is the cornerstone of the generation of perceptually separated color classes. It is also possible to compute color distance with the Deltae method, but this is not sufficient. We also want colors with everyday language semantics (this constraint will help users to designate colors and then the routes orally). Constant value of luminosity and hue do not automatically produce colors with a correspondence in the

language. To address this issue, we based the extraction of color with the "named colors" experimentations [3] [17]. Finally, we use the Munsell tables [14] to extract, with a given luminosity, the available Hue in the « named color » space.

```

createRGBColorset (numColors, colorModel, ICCprofil, startAngle,extend)
{
    colorModel = 'LCHab' if not defined
    ICCprofil = 'sRGB' if not defined
    startAngle = 0.0 if not defined
    extend = 360.0 if not defined
    stepAngle = extend / numColors
    RGBcolors = array of numColors color
    foreach i in {1 .. numColors}
    {
        newH = startAngle + stepAngle * i
        newL = namedColorLum(newH,colorModel)
        newC =
maxChromaInRGBgamut(newH,newL,colorModel,ICCprofil)
        RGBcolorsi = LCHcolor2RGB(newL,newC,newH,
        colorModel,ICCprofil)
    }
    return RGBcolors
}

```

The pseudo code presents the extraction of an RGB color set with the number of colors to produce, the color model to use (by default LCHab), the ICC profile of the output device (by default sRGB), the starting angle of the Hue, and the available range. The function **namedColorLum** returns the best luminosity value for a given hue. The function **maxChromaInRGBgamut** computes the maximum Chroma (i.e., saturation) in the gamut. Finally, the function **LCHcolor2RGB** converts the RGB value of an LCH color with the ICC profile.

This algorithm could produce a countless number of colors, but in practice, a sector contains of maximum of 30 FRs and their corresponding color.

A. Colors assignments

With a metro map composed of N routes, we can generate, with our color generation algorithm, N visually separable colors. But $N!$ available choices do exist to assign them to the N routes. We need to set up a numerical value that corresponds to a quality level of this color assignment. We want to assign to close routes the most distinguishable color. To compute the color distance we use the Delta_e 2000 formula.

In order to compute the distance between lines, we consider lines as discreet points regularly spaced:

$$\text{i.e. } l_1 = \{ Pt_{11}, \dots, Pt_{1m} \}.$$

The standard distance between two lines is the minimum distance that separates two points of each line. This distance is not suitable regarding the visual perception of distance. For instance, two crossing lines have, with this distance computation, a distance equal to zero, whereas perceptually they are very distinct. Therefore, we define a new distance computation which is close to the perceptual distance between lines:

$$l_1 = \{ Pt_{11}, \dots, Pt_{1m} \} \text{ et } l_2 = \{ Pt_{21}, \dots, Pt_{2q} \} \text{ with } m < q$$

$$\text{dist } (l_1, l_2) = \frac{(\sum_{i=1}^m \min(d(Pt_{1i}, Pt_{2j}), j = 1 \rightarrow q))}{m}$$

In this formula, d is the Euclidean distance between two points. Thus, the distance between the line l1 and l2 is the sum of the minimum distances between each point of l2 and each point of l1. Finally, we normalize this distance by dividing it by the biggest distance between the lines.

A color assignment is a couple of a line and color couple (l_i, C_i). We define the quality of a color assignment as following:

$$\text{qual} = \sum_{i,j,i < j}^N \text{Delta}_e(C_i, C_j) \times 2 \times (1 - \frac{1}{1 + e^{-\text{dist}(l_i, l_j)}})$$

This value corresponds to a weighting mean of Δ_e .

The weight of each Δ_e depends of the distance between lines. We apply a sigmoid curve to this distance in such a way that its weight tends towards 1 when the distance tends toward 0 (the case of two close lines), and towards 0 when the distance tends towards 1 (case of distant lines). Thus, the weight of the Δ_e increases when the distance between lines decreases.

Finding the optimal assignment between the $N!$ assignments is an NP complete problem. No deterministic methods can solve it in a reasonable time span. We will therefore use the simulated annealing method which is very efficient for this kind of problem. In our case, we try to optimize the color assignment. Colors are computed beforehand. The current state is a group of couples color, line, a neighbor is computed by switching the color of two couples. Cooling parameters of simulated annealing are tuned to 0.9997. The algorithm stops when it converges, i.e. 100 steps without increasing the function.

B. Optimal or worst color assignment

The optimal view is the **Figure 8**: it corresponds to optimized geographical positions and an optimized crossing with an optimized color assignment.

We also produced maps with the worst color assignment. Therefore we use the above optimization algorithm with the opposite of the quality function. Thus, the view produced displays close lines with close perceptual colors.

We observe, especially in **Figure 8**, that the opposite of the cost function produces assignments where close FRs have perceptually close colors. This creates a rainbow effect through FRs. On the contrary, with the original cost function, close FRs are discriminated by colors as expected.

In the next section we assess the views produced.

VI. EXPERIMENTS

Firstly, we performed a qualitative evaluation by showing the produced metro-maps produced to Air Traffic Controllers. We assessed the extent to which the corresponding sector was easier to recognize and if the view produced was clearer than the original one. As a first result, the optimized geographical

views are clearer than the original view and are easier to recognize.

Secondly, we performed a quantitative evaluation to assess the quality of the metro maps produced.

In this experiment, we selected nineteen participants without sight problems (color blind). They were seated facing an 18" video monitor located at a distance of 0.80 m from their head.

The protocol simulates a real use case with a FR metro-like map: users have to accurately select a specific section of a Flight Route as fast as possible.

Two different color assignments have been tested:

- Optimal color assignment: close FR have distant color,
- Worst color assignment: close FR have close color.

We justify our worst color choice (i.e. non discriminating color assignment) as compared to a random color placement, because random can, with serendipity, create a good color assignment (strong discriminating power).

Two different geographical topologies have been tested:

- No geographical optimization, which corresponds to the initial view without beacon spacing and with tangled FR,
- Geographical optimization, which corresponds to the optimized view, with beacon spacing, and FR crossing optimization (few crossing FR).

We created metro maps of five ATC sectors. Each sector produces four different metro maps:

- No geographical optimization with the worst color assignment.
- No geographical optimization with the best color assignment.
- Geographical optimization with the worst color assignment.
- Geographical optimization with the best color assignment.

We also defined three difficulties levels for each section of FR:

- Level 1 (easy): the section is located in a non dense area without crossing lines,
- Level 2 (medium): the section is located in a dense area with less than 2 crossing sections,
- Level 3 (hard): the section is located in a dense area with more than 2 crossing lines.

Then, for each map with extracted 3 sections of each difficulties level (easy, medium, and hard) which represents 9 sections per map.

Each participant had to select 45 sections (5 maps, 9 sections per map) with two kinds of metro map:

- The worst color assignment with, and without, geographical optimization,
- Or
- The best color assignment with, and without, geographical optimization.

Each participant performed 90 randomized tries (45 sections with two conditions). The number of maps and the kinds of map (worst colors or optimized color assignment) reduced the learning effect. We recorded the time needed to select a section of a FR (this time record starts when the requested color and the blinking beacons appear) and the correctness of the answer (we counted the number of right or wrong answers).

A. Procedure

In the view produced, a horizontal straight line (at the top left corner of the view) indicates the requested color (Figure 2). Two blinking beacons indicate the requested section (Figure 2). When the user flies over a section of a FR with the mouse pointer, the fly-over section oscillates indicating that this section can be selected by pressing the mouse (Figure 2).

For each try, the mouse pointer is located over the requested color position (the horizontal straight line); hence the user can memorize the color. The two blinking beacons use the pe-attentive effect [21], so the user immediately sees which area must be investigated. Finally, the user has to select the correct section by choosing the requested color between the blinking beacons. The fly-over section is animated; other visual designs were envisaged but they created color distortions: wider line sections, blinking colors.

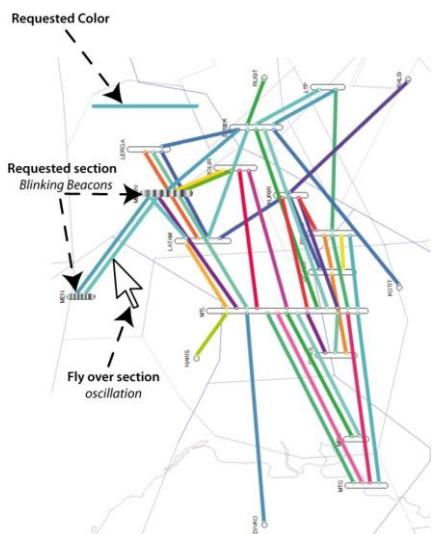


Figure 2 : The User's task is to select the section of a Flight Route between two blinking beacons with a requested color.

B. Results

All dependent measures were separately analyzed by performing an analysis of variance (ANOVA) with two groups of participants (no geographical optimization, and geographical optimization) and with repeated measurements for two factors: levels of difficulty (easy, medium, and hard) and color assignments (best and worst color).

Concerning the selection time, there was a significant effect regarding the level of difficulty $F(2,28)= 45.77 p <.05$; the greater the level of difficulty , the longer the time movement (Figure 3).

In addition, there is an interaction between the level of difficulty and the geographical condition (with or without geographical optimization), $F(2,28)= 4.33, p <.05$ (Figure 3). The time movement increase, due to the level of difficulty, is less important with the geographical optimized view than with the non geographical optimized one (Fig. 4). Therefore, the geographical optimized view is more efficient (smaller selection time) than the non geographical optimized one.

The analysis of the error number yielded a main effect with the level of difficulty $F(2,28)= 8.79 p <.05$: The number of errors increases when the level of difficulty increases.

In addition, there is an interaction between the level of difficulty and the color assignment conditions (best or worst), $F(2,28)= 40.79 p <.05$ (Figure 4). The error number increase, due to the level of difficulty, is less with the best color assignment than with the worst one (Figure 4). Therefore, the good color assigned view is more efficient (smaller errors number) than the worst color assigned one.

As expected, the two proposed enhancements (geographical optimization and best color assignments) improve performance in the select a section of a FR.

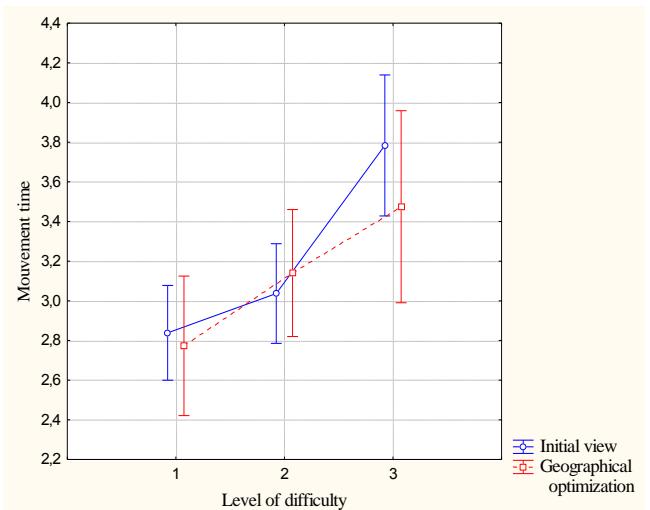


Figure 3 : Movement time as a function of the level of difficulty and geographical conditions.

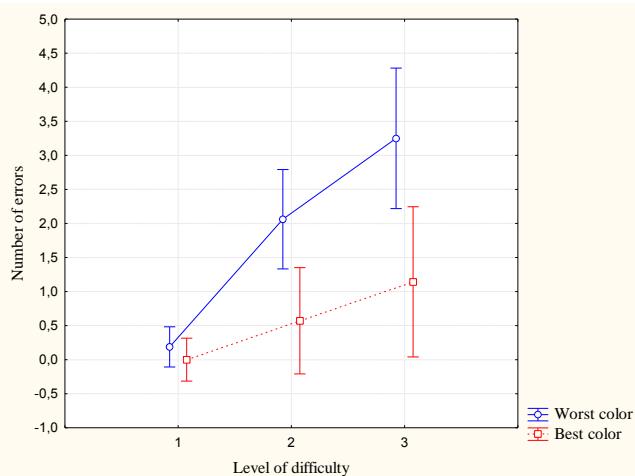


Figure 4 : Number of errors in terms of the level of difficulty and color conditions.

VII. CONCLUSION

In this paper, we address a new issue of metro map layout. We present a complete method to produce an efficient layout of aircraft Flight Routes in the ATC context. We turn ATC specific visual constraints into mathematical formulations. The simulated annealing algorithm, with these adapted cost functions, and optimizations, produces visualizations which fulfil the defined constraints. As we use an automatic process, the method allows the generation of metro-like views for any ATC sector. The automatic generation process is a great asset when sectors or FRs are updated.

We also studied the generation and the placement of colors on metro maps in order to enhance the readability of the view. This issue hasn't been extensively studied until now. We proposed a method to generate colors that takes into account their semantics and their perceptual distance with respect to other colors. We defined a cost function that measures the quality of a placement of color on a metro map. This cost function is integrated into an optimization process in order to obtain an optimal view.

We validated the produced views with qualitative and quantitative assessments. The geographical optimizations with optimized color views are clearer than the original view, and easier to recognize. The views produced improve the accuracy and speed up the selection of a specific section of a FR.

The proposed method is not specific to the ATC context and can be used for any metro- like visualization.

We plan to use the produced views produced with Air Traffic Controllers. Two main applications are envisaged. Firstly, we plan to add the traffic density of each FR the view. As a design choice, we plan to use the width of the line or its luminosity. Secondly, we will conduct a design study to use these maps to help companies to reserve time-slots on FR.

REFERENCES

- [1] Agrawala, M. and Stolte, C. 2001. Rendering effective route maps: improving usability through generalization. In Proceedings of the 28th Annual Conference on Computer Graphics and interactive Techniques SIGGRAPH '01. ACM, New York, NY, 241-249.
- [2] Bekos M., Kaufmann M., Potika K., Symvonis A. : Line Crossing Minimization on Metro Maps. Proc. of Graph Drawing 2007. 231-242.
- [3] Berlin, B., Kay, P., Basic color terms: their uni-versality and evolution. Berkeley; Oxford: University of California Press (1969) pg. 196.
- [4] A Color Appearance Model for Colour Management Systems: CIE CAM 2002, CIE 159, (2004)
- [5] Denis M.: The description of routes: A cognitive approach to the production of spatial discourse. Cahiers de Psychologie Cognitive, 1997.
- [6] Garland K.: Mr Beck's Underground Map, Capital Transportation Publishing, 1994.
- [7] Ishizaki, S. 1995. Color adaptive graphics: what you see in your color palette isn't what you get! In Conference Companion on Human Factors in Computing Systems (Denver, Colorado, United States, May 07 - 11, 1995). I. Katz, R. Mack, and L. Marks, Eds. CHI '95. ACM, New York, NY, 300-301.
- [8] Harrower, M. and Brewer, C.A., ColorBrewer.org: an online tool for selecting colour schemes for maps. Cartogr. J. v40 i1. 27-37.
- [9] Kirkpatrick S., Gelatt C. D., Vecchi M. P.: Optimization by simulated annealing. Science 220, 671-680, 1983.
- [10] Levkowitz, H. and Herman, G. T. 1992. Color Scales for Image Data. IEEE Comput. Graph. Appl. 12, 1 (Jan. 1992), 72-80.
- [11] Mackay W., Fayard A-L, Frobert, L., Médini L. : Reinventing the Familiar: Exploring an Augmented Reality Design Space for Air Traffic Control. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 1998. 558-565.
- [12] Merrick D., Gudmundsson J.: Path Simplification for Metro Map Layout. Proc in Graph Drawing 2006. 258-269.
- [13] Moretti, G.S. and Lyons, P.J. Controlling the complexity of grouped items in colour interfaces. Proceedings of the 6th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction: making CHI natural, ACM (2005), 19-23.
- [14] Munsell, A. H. (1912). "A Pigment color System and Notation". The American Journal of Psychology.
- [15] Nesbitt, K.V.: Getting to more abstract places using the metro map metaphor. Proceedings of the Information Visualisation International Conference, IV 2004, IEEE Computer Society, pp488-493, 2004.
- [16] Nöllenburg, N. and Wolf, A.: A mixed-integer program for drawing high-quality metro maps. Proc. 13th Internat. Sympos. Graph Drawing (GD'05), volume 3843 of Lecture Notes in Computer Science, pages 321-333. Springer-Verlag, 2006.
- [17] Párraga, A. Benavente, R Vanrell, M. and Baldrich, R. Modelling inter-colour regions of Colour Naming Space. In Proceedings on Color in Graphics, Imaging and Vision (CGIV'08).
- [18] Rheingans, P., Task-based Color Scale Design. Proceedings of Applied Image and Pattern Recognition '99, SPIE, pp. 35-43.
- [19] Robertson, P. and O'Callaghan, J. 1986. The Generation of Color Sequences for Univariate and Bivariate Mapping. IEEE Comput. Graph. Appl. 6, 2 (Feb. 1986), 24-32.
- [20] Stott, J. M. and Rodgers, P.: Metro Map Layout Using Multicriteria Optimization. In Proceedings of the Information Visualisation International Conference. IV. IEEE Computer Society, pp355-362, 2004.
- [21] Treisman, A., Preattentive Processing in Vision, Computer Vision, Graphics, and Image Processing, 31(2):156-177, August 1985.
- [22] Tufte, E. R. Visual Explanation. Graphics press, Cheshire 1997.
- [23] Ware C., Information Visualization, perception for design, Morgan Kaufmann, 2002.
- [24] Ware, C. 1988. Color Sequences for Univariate Maps: Theory, Experiments and Principles. IEEE Comput. Graph. Appl. 8, 5 (Sep. 1988), 41-49.

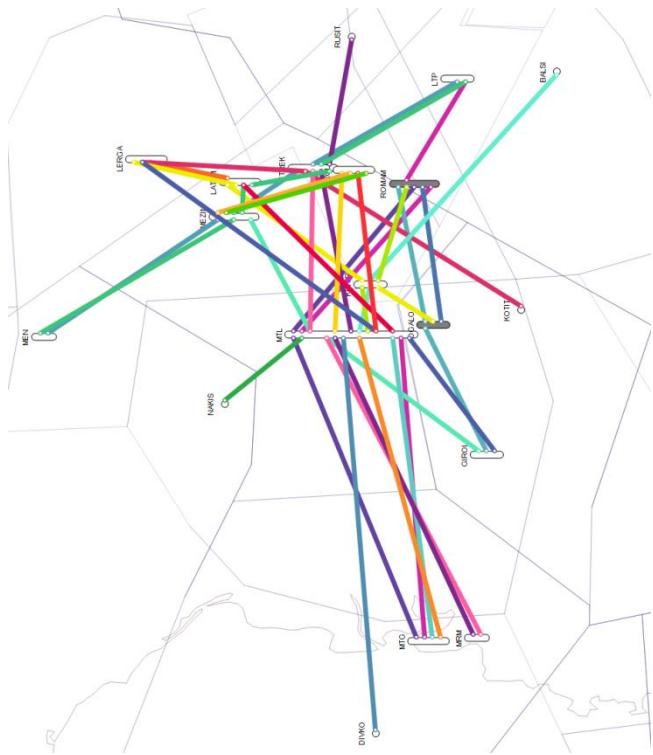


Figure 5 : View with no geographical optimization and with the best color assignment.

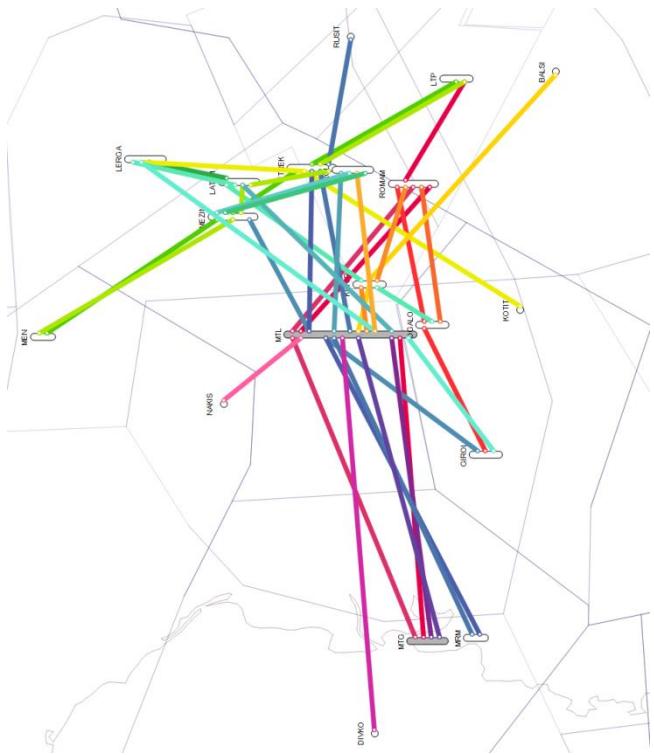


Figure 7 : View with no geographical optimization and with the worst color assignment. This view is the less efficient one.

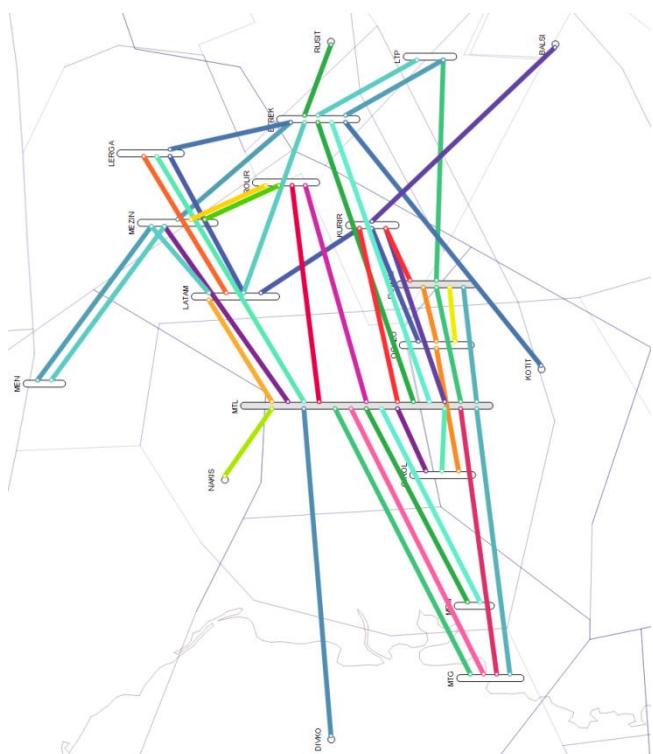


Figure 6 : View with geographical optimization and with the best color assignment. This view is the most efficient one.

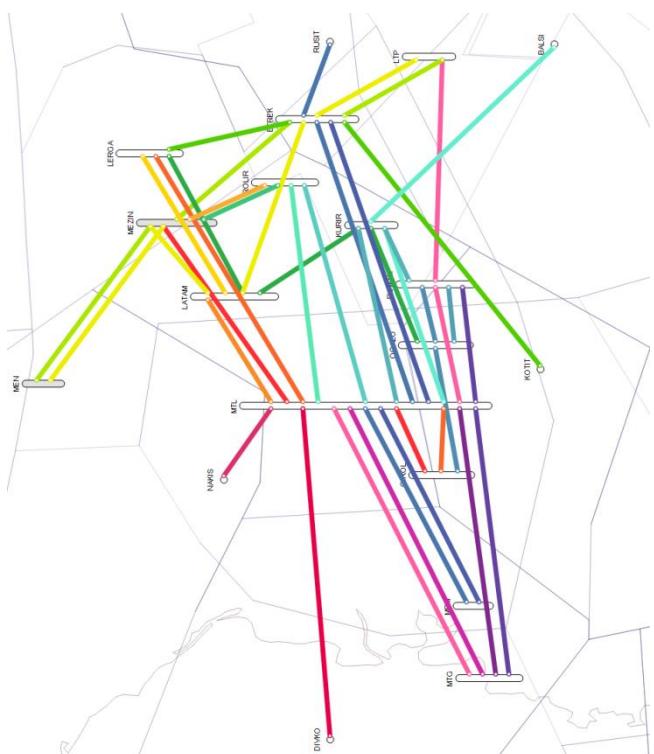


Figure 8 : View with geographical optimization and with the worst color assignment.

A Participatory Design for the Visualization of Airspace Configuration Forecasts

Nicolas Saporito
DSNA/DTI R&D
7, avenue Edouard. Belin
31055 Toulouse France
Nicolas.Saporito@aviation-civile.gouv.fr

Christophe HURTER
DSNA/DTI R&D
IRIT
7, avenue Edouard. Belin
31055 Toulouse France
christophe.hurter@aviation-civile.gouv.fr

David GIANAZZA
DSNA/DTI R&D
IRIT
7, avenue Edouard. Belin
31055 Toulouse France
David.Gianazza@aviation-civile.gouv.fr

Geraldine BEBOUX
DSNA/DTI R&D
7, avenue Edouard Belin
31055 Toulouse France
Geraldine.Beboux@aviation-civile.gouv.fr

Abstract— Currently, airspace-related activities in Air Traffic Control Centers (ATCC) are dispatched between the Flow Management Position (FMP) operators and the control room manager, and take place in two different time frames. The first activity (FMP) is the planning, 2 days ahead, of airspace usage and anticipated overloads, using coarse-grain and relatively inaccurate workload prediction metrics. The second activity (control room manager) is the day-to-day operation, where workload is re-assessed in real-time and where airspace may be re-configured according to the actual traffic of the day. In previous works, a workload model relying on relevant air traffic complexity metrics was proposed, using a neural network trained on past sector operations. This workload prediction model was combined with tree search algorithms, in order to compute optimal partitions of the airspace in Air Traffic Control (ATC) sectors. This method provides more accurate airspace configuration forecasts than today, thus improving the overall predictability of the Air Traffic Management (ATM)/ATC system. When relying on accurate 4D-trajectory predictions, as expected in the SESAR program, it could contribute towards bridging the current gap between the pre-tactical airspace/flow management and real-time operations. In this paper, we detail the participatory design approach that we used to develop a research prototype displaying the algorithm's results. As there is no such forecasting tool today, the main issue was to create a user interface in the absence of an existing user.

Keywords: User-Centered Design, Human Computer Interaction, Airspace Configuration Forecasts

I. INTRODUCTION

With the global trend towards increasing traffic over the last few decades, research issues in Air Traffic Management (ATM) have become more and more critical in the development of future concepts and systems. However, although research is very active in the ATM field, the implementation of its outputs in the form of real-life software, that can actually be used in an operational context, is a fairly slow process. It has not always met all the great expectations that have, in the past, been placed on its ability to enable the ATM system to handle safely and efficiently an increasing amount of traffic.

As a consequence, it could be assumed that research in ATM is disappointing in terms of results. The stakeholders and institutions funding the ATM R&D activity are certainly tempted to re-organize the research activities into a standard

industrial V-cycle, so as to lead research towards a common goal: the efficient development of the next generation of ATM systems. This ambitious objective is legitimate, and in this context the V-model might be useful to develop high quality industrial software when the technology is mature and the users' needs are well known.

In terms of research however, this approach is clearly counter-productive and costly when addressing ill-posed problems like: "how could we improve the current ATM system with a new operational concept?" This is a rather ambiguous question that may have many context-dependent answers. Research is usually much more efficient when addressing open questions which seek non-ambiguous answers. It is even more efficient when several teams can work in parallel on a common well-posed problem, exploring different paths with various methods, in collaboration, and/or in competition, with each other. But this discussion is not within the scope of our paper.

In the standard V-cycle model of development, the users are mainly involved at the beginning (definition of their needs) and at the end of the cycle (validation of the final product). If the initial operational concept proves impractical in real-life, or if the derived software and systems do not ultimately meet their objectives, the whole cycle has to be repeated: a new concept, new research, new software developments, and so on. Furthermore, this development model may not be suitable when validating new ideas and algorithms in a context that may change, depending on research results and on the users' feedback. Rapid iterative development involving the users and the researchers at several intermediate steps is to be preferred in this case, when developing such research prototypes.

In this paper, we used a participatory user-centered method to develop an interface for a new algorithm providing airspace configuration forecasts. This new algorithm combines tree-search methods with a neural network, assessing the air traffic controllers' workload, in order to compute optimal partitions of the airspace in Air Traffic Control (ATC) sectors. The neural network was trained on past sector operations, considering existing ATC sectors that were split, merged, or recombined according to the actual workload. The Graphic Interface presented in this paper is a research prototype that displays the algorithm's results. It aims at demonstrating and improving these algorithms, taking into account the feedback of potential users so as to provide realistic forecasts.

The problem being addressed is well-posed in the sense that an optimal airspace configuration can be computed from a given traffic prediction, and the realism of the resulting configuration can be improved by considering well-defined rules when partitioning the airspace. However, the operational context in which such a forecasting tool could be used is not clearly defined. In the current European system, airspace and workload management activities are dispatched between the Flow Management Position (FMP) operator, who makes coarse-grain and approximate pre-tactical forecasts one or two days ahead, and the control room manager who decides to split, merge, or recombine ATC sectors in real-time, according to the effective workload of the Air Traffic Controllers (ATCOs). Forecasting algorithms able to provide more accurate predictions a few hours ahead could certainly be used somewhere between the current FMP pre-tactical prediction and real time traffic control, but such a tool does not exist today for airspace management purposes.

In other words, there is not yet an operational concept, and no “final user” to interview when designing the Graphic interface of our research prototype. The study of user tasks and the realization of an appropriate interface is already a difficult task when the user, his or her role and activities, are clearly identified. In our case, the lack of a final user was an additional difficulty in the design process and we had to ask “potential users” (operational experts, ATCOs, including a control room manager) to imagine the operational use of the proposed algorithms. The existence of these issues was the reason for the use of a participatory user-centered design process in which potential users, researchers, and Human Computer Interaction (HCI) experts are involved throughout the design of the Graphic Interface, instead of standard design and development methods, in which the users are mainly involved at the beginning and at the end of the process.

The main content of this paper is organized as follows. Firstly, we describe how the airspace management activities take place in the current operational context, and how they could be envisaged in the future SESAR operational concept. Secondly, we describe briefly the new algorithms forecasting the ATCOs workload and the airspace configuration, focusing on the outputs that may be significant to the final user. Thirdly, we lay out the principles of the participatory user-centered design. Fourthly, we describe the results of the application of this iterative process in terms of helping to solve our problem, and the current version of the resulting Graphic Interface. Finally, the main issues and results are summarized in the concluding section, together with the perspectives of future developments in the context of three SESAR projects.

II. CONTEXT

Currently, the Flow Management Position (FMP) located in the Air Traffic Control Center (ATCC), in collaboration with the Central Flow Management Unit in Brussels (CFMU), organizes traffic management. The traffic management on a given day is sequenced with several steps (Figure 1).

Several months beforehand, at the strategic level, the forecast traffic and sector capacities are analyzed to detect

potential anomalies, and the strategic airspace design and flow orientation schemes are amended.

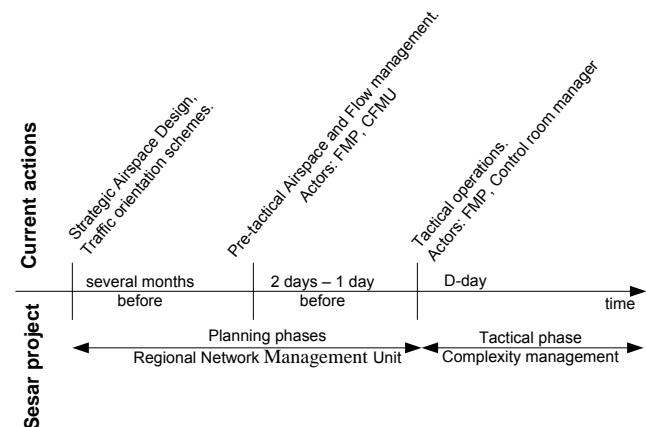


Figure 1: Current and SESAR program [13] timeline of actions for airspace and flow management.

Two days beforehand (pre-tactical level), the FMP checks the controllers' duty roster and produces a sector opening schedule based on the predicted traffic and sector capacities. To do so, the traffic demand is evaluated from two data sources:

- FMP can access the previous traffic record up to seven days before the requested period,
- FMP can access a traffic prediction provided by the CFMU.

The current method used by the operator to forecast the airspace configuration (the sector opening schedule) is fairly simple. Considering a set of static airspace partitions, that have been previously filed in the system by the FMP staff, the operator selects, for each time step (usually 30 or 60 minutes) the best configuration among those he or she thinks may be the most adequate for this time of the day. The prospective airspace partitions are evaluated by comparing the traffic flow entering each ATC sector to a given sector-specific threshold (the sector capacity). This choice of the best configuration is subjective and based on the past experience of the operator. Once this sector opening schedule is built, there may remain some ATC sectors for which the predicted incoming flow is higher than the capacity. Such potential overload problems are signaled to the CFMU which may enforce some regulations on traffic flows contributing to such overloads.

One day beforehand, the CFMU publishes restrictions on the Flight Plan to the Air Traffic Services and Operators (Airline companies). The FMP defines the ATC capacity according to the exact number of available ATCOs and defines the number of sectors that can be opened.

On D-day, the FMP tries to optimize relations between demand and capacity, to reduce delays in collaboration with the ATC Supervisor and CFMU. Comparing, hour by hour, the traffic demand to the sector capacities, FMP adjusts necessary traffic restrictions. To do so, the FMP operator can perform different actions:

- He anticipates sector overloads, negotiates traffic regulations with the CFMU,
- He helps the supervisor in the splitting/merging management of sectors,
- He answers controllers' requests concerning regulations required, or available re-routings,
- He reduces, in collaboration with CFMU, the delays generated by regulations required within the ATCC or adjacent ATCCs.

In the SESAR program, practitioners' actions are being redefined in order to deal with traffic complexity and density (Figure 1) [13]. The following corresponds to our understanding of the different roles of practitioners in airspace management and may change with the future evolutions of the project.

During the Planning Phase, the Regional Network Management Unit, a kind of FMP, will match overall capacity to demand. In real time, the Complexity Management will optimize the airspace configuration and traffic flows in order to keep the traffic complexity at an acceptable level for ATCOs (p 30 in [13]). The Regional Network Management Unit and the Complexity Manager may both use multi-sector tools.

III. PARTICIPATORY DESIGN APPROACH

In our case, the user tasks are not clearly defined. Therefore, the standard User-Centered Design (UCD) is not fully satisfying. Observations and interviews at the beginning of the design process are not sufficient to define accurately what the user's tasks will be, or could be, in a future context different from today's operations. Participatory design involves the users all along the design process. It is an ongoing research objective [14] and has already been used in the Air Traffic Control field to experiment the augmented paper strip [11] (a mix of paper and informatics strip).

Basically, participatory design assumes that users know what they need and can have innovative ideas [8]. Nevertheless, users are confronted with the difficulty of expressing their needs and finding out how to address them. The approach here is two-ways (as opposed to the unidirectional approach of User-Centered Design): the user is not only observed and interviewed (as in standard UCD), but also integrated into an iterative design process [11], where he is helped by the designers to express his needs clearly and is also repeatedly involved in validation exercises. This ensures that the design produced at each step of the process meets the users' needs and is actually usable.

In our project, the final user of the interface is not clearly identified. However, existing operators in charge of airspace-related tasks are well identified: FMP operators and control room managers. Therefore we involved them in our design process with the help of computer scientists (the designer of the algorithm and specialists in Human/Computer Interaction). We organized the reflection around brainstorming and workshops using participatory design methods. We first thought about what kind of improvement our tool could bring to the tasks of the current FMP. We then tried to imagine how to help a future

operator who would dynamically reconfigure airspace at the tactical level in the context of the future ATM systems envisaged by the SESAR program.

IV. AIRSPACE CONFIGURATION FORECASTING

The algorithms forecasting the airspace configuration have already been presented in detail in past publications ([2], [3], [4]), as well as the selection of relevant complexity metrics used as input to these algorithms ([5], [6], [7]). The results presented in these publications show that the computed output is fairly close to the number of ATC sectors that were actually operated. Further work on the algorithms will mainly deal with the introduction of constraints on the transitions between successive configurations, so as to get closer to the way sectors are actually split, merged, or recombined in the field.

In this section, we will very briefly present the hybrid method that was used to forecast airspace configurations, mainly focusing on the features useful to the design of the user-interface. Considering that the airspace is divided into several airspace modules¹, we are looking for the optimal partition of the airspace into ATC sectors (each sector is made up of one or several modules) that may be operated by controllers under normal workload conditions.

To that purpose, a branch & bound algorithm is used to explore all possible partitions. This tree search algorithm is combined with a neural network assessing the controllers' workload for each ATC sector. The neural network is trained on past sector operations, using the fact that ATC sectors are usually split into several smaller sectors when the workload is excessive, or merged with other sectors when the workload is low.

The dynamic behavior of the algorithm is the following: we start from an initial airspace configuration at time t_0 and consider a time interval $[t_0; t_1]$ in which we want to forecast the next optimal airspace partitions. The workload in each ATC sector of the current configuration is checked at each time step $t + \delta t$ in the chosen time interval, using the neural network. If the probabilities computed by the neural network show that the workload is either too high or too low in one or several sectors, an airspace re-partitioning is triggered. In that case, the branch & bound computes a new optimal configuration. It minimizes cost in terms of the number of ATC sectors and the workload in each sector. The neural network is once again used to assess the workload in the sectors of the prospective configurations.

The airspace partitioning may also take into account other constraints, such as the maximum number of controller working positions available throughout the day.

The neural network may be seen as a statistical model tuned so as to minimize the error between the computed output and some observed data. We used a simple feed-forward network (see [1] and [12] for more details on the theory and algorithms of the neural networks) with 6 input units, 15 units in the

¹ These modules are usually called airspace sectors, but we shall denote them as airspace modules in the rest of this paper, so as to avoid confusion with the air traffic control (ATC) sectors operated by air traffic controllers, which are made up of one or several modules.

hidden layer, and 3 output units, with the following equation where the weights ω were tuned on recorded data:

$$(y_0, y_1, y_2)^T = \Psi \left(\sum_{j=1}^{15} w_{jk} \Phi \left(\sum_{i=1}^6 w_{ij} x_i + w_{0j} \right) + w_{0k} \right)$$

where Ψ is the softmax function: $\Psi_k(z_k) = \frac{e^{z_k}}{\sum_{m=1}^3 e^{z_m}}$ and Φ is the sigmoid logistic function: $\Phi(z) = \frac{1}{1+e^{-z}}$

The input variables $x = (x_1, \dots, x_6)$ are the complexity metrics that were found to be the most relevant for our purpose ([5], [6]), normalized by subtracting the mean value and dividing by the standard deviation, and smoothed using a moving average method ([7]). These metrics are the sector volume V , the number of aircraft within the sector N_b , the average vertical speed avg_vs , the incoming flows with time horizons of 15 minutes and 60 minutes ($F15, F60$), and the number of potential trajectory crossings with an angle greater than 20 degrees ($inter_hori$).

The output vector $y = (y_0, y_1, y_2)$ can be interpreted as a vector of posterior probabilities of class-membership ([10]): y_0 can be seen as the probability $p(C_{low}=x)$ that the ATC sector falls in the "low workload" class when the measured air traffic complexity vector is x , and similarly for y_1 and y_2 , when the classes are C_{normal} and C_{high} respectively. Using an abbreviated notation, we shall denote as $y = (p_{low}, p_{normal}, p_{high})^T$ the output vector in the rest of this paper, so as to clarify the nature of the neural network output.

As we are necessarily in one of the above three cases (low, normal, or excessive workload), the sum of the three probabilities p_{low} , p_{normal} , and p_{high} is invariably 1 (This is ensured by the use of the softmax function in the output layer).

This allows the three probabilities issued by the neural networks to be displayed, stacked one above the other, in a box of height 1, as shown in Figure 2, where the following color code was chosen: blue for p_{low} , green for p_{normal} , and red for p_{high} . The red line shows the chosen threshold for the p_{high} probability, above which a reconfiguration is triggered. Similarly, the blue line is the threshold value for probability p_{low} .

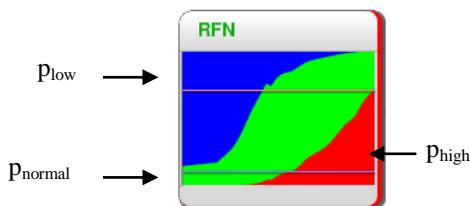


Figure 2: Three probabilities: p_{low} , p_{normal} , and p_{high}

To conclude this section, let us summarize the output data that may be available to the final user. The final output of the forecasting algorithm is a sequence of airspace configurations with their time intervals. Each configuration is a list of ATC sectors for which a workload indication (the neural network output probabilities) is available. There are two threshold values for the low and high workload probabilities. These are parameters of the decision criteria used to trigger the airspace configuration changes. The transitions between successive

configurations are also computed, and may help users to track how the airspace modules are recombined.

V. ITERATIONS: SEVERAL WORK SESSIONS WITH POTENTIAL USERS

This section illustrates our participatory design process with the details of the different steps.

A. 1st session: context and presentation of the algorithm

In this first session, we presented the airspace partitioning algorithm. Most of the discussion was devoted to how this algorithm could be used in the current airspace and flow management context (FMP, CFMU, tactical airspace configuration) or in the future SESAR operational concept (complexity management, multi-sector planning). A group of 6 participants was set up: the designer of the algorithms, HCI specialists, former control room managers and FMP operators.

The session lasted 1h30 and was divided into three parts. After two short lectures (around 15 min each) introducing the algorithm, the FMP tasks today, and the SESAR concept, the group was invited to participate in a brainstorming session (30min) on the following theme: what kind of data will the operator(s) need? The ideas generated during the brainstorming were then discussed during the last 30 minutes. As a result, the participants identified the most relevant outputs provided by the algorithm:

- the computed airspace configurations,
- the transitions between consecutive configurations,
- the workload prediction for each ATC sector, in the form of probability indicators (low/high/normal).

The workload probabilities were considered sufficient for the moment. A further investigation will be conducted to select other potential complexity metrics from among the data used by the algorithm.

The users also expressed the need to identify clearly the events that triggered a configuration change, and the durable overloads that may lead to dangerous situations.

B. 2nd session: brainstorming

The same group participated in a second session. The goal was to be more specific and to find practical ideas for the Graphic User Interface, both for the current task of the FMP and the future task of the Multi Sector Planner (MSP).

There were four steps during the two hours of this session. The first 20 minutes were devoted to a recap of the previous session. Then, secondly, the aims and the practical details of the participatory design process were detailed. Thirdly, the participants were involved in a brainstorming session of around 50min on the following themes: How to display the workload evolution of the ATC sectors (information of low, normal and high workload)? and How to navigate in this display (zoom, translations...)? The proposed ideas were then sorted for 10 minutes. Finally, the last step was a design walk-through phase of 40 minutes, which consisted of "quick and dirty"

prototyping of the selected ideas with paper, pencils and scissors.

Figures 3 and 4 show some prototype results of the design walk-through. They give an overview of the visualization imagined by the users. Figure 3 shows a compact view with only the names of the ATC sectors with a color code for their status and their potential durable overload. Figure 4 shows the result of a click on a configuration (that at t=264s). The configuration and its surrounding configurations are unfolded to show the relationship between each of the sectors composing them (referred to as inheritance later in this paper) and the detailed workloads. The folding/unfolding paradigm, which only displays a part of the available information, will be further developed in the description of the interface established.

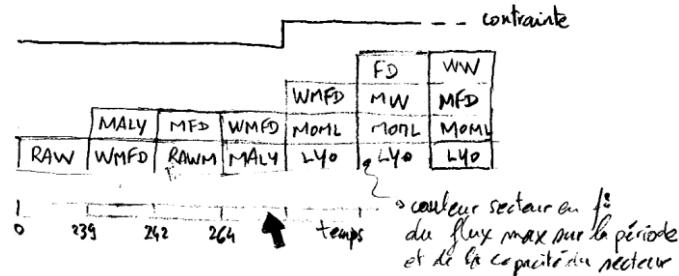


Figure 3: Draft of the overview visualization

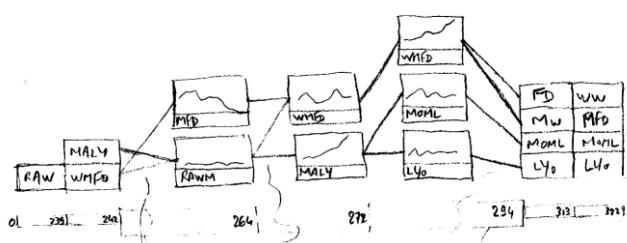


Figure 4: Draft of the sector details

C. Development of the prototype

The next phase of the project was the first iteration of the software prototype development. At this level, the prototype basically presents the results of the algorithm: the forecast sector workload, the best configuration propositions, and the merge/split/recombine events over the day, according to the suggestions from the first sessions. Currently, these data are computed "off-line" for a whole day of traffic. Typically no interaction to accept/refuse the proposed configurations is available. However, the user can interact with this visualization to explore and understand the sequence of events thus assessing the algorithm's behavior.

This is the first development iteration of this research prototype. Future iterations should allow the user to compute the airspace configuration "on-line" through the interface. An additional feature would be to forecast dynamically the configuration on a receding time horizon.

D. 3rd session: evaluation of the prototype

The third session consisted of the evaluation of the prototype with the same persons (plus one HCI specialist). This session lasted two hours with a short recap of the situation (10 min), followed by a plenary presentation of the prototype (10 min).

The group was then split into two skills-oriented subgroups (HCI specialists on one side, ATC specialists on the other) in two different rooms to participate in a handling session. For 40 min, the participants became familiar with the prototype by manipulating it under the supervision of the designers. Questions were answered by the supervisors and remarks/suggestions were noted.

The two subgroups were then reunited for a plenary discussion of one hour of which the goals were to evaluate the first brainstorming choices and assess the prototype as a pre-practical tool for helping the existing FMP; it also aimed to widen the scope of our vision, in order to consider the possibility of a tactical tool for a yet-to-be-invented user. All the brainstorming remarks were compiled and discussed by the whole group. The points of view of the different participants were confronted and, depending on the HCI/operational origin, some of the suggestions made by a subgroup were developed or modified by the other subgroup, particularly concerning the information displayed in the global view, as we will see in the next section.

There were two kinds of suggestions. Some of the remarks aimed at improving the interface and the interactions; others were proposals for additional interactions for the evolution of the prototype towards a tactical tool: acceptance/refusal of the proposed configurations, manual reorganization of configurations, a what-if function (what will the sector workloads be if I reorganize the airspace like this?)...

VI. CURRENT PROTOTYPE

In this section, we will describe the interface that was produced taking into account the results of the third session (evaluation of the first prototype).

A. Description of the interface

The quantity of information to be displayed is sizeable: successive airspace configurations over the day, transitions between configurations, workload prediction for each control sector at every minute of the day, or other complexity metrics on demand. Consequently, to avoid confusing the user with a view of excessive complexity, we need to present both a global general view and a detailed zoomed one. The user must be able to switch quickly from one to the other, and, moreover, must not lose the focus on the general trend when in a detailed "mode". Therefore it was initially decided to produce a flexible representation of the day, rather than two distinct modes (general/zoomed). A fish-eye [9] type display would have presented all the configurations of the day with a low level of detail and, on demand, would have been able to "unfold" configurations in order to present more detailed information on smaller zones while keeping the schematic and global representation of the rest of the display.

But the standard display size didn't allow displaying all the information even if all the configurations were folded. Users requested an abstract view with the summary of relevant events during a day:

- Number of sectors for each configuration (represented as stacks of sectors),
- Global workload for each sector. A color that fills the box indicates the sector state (low, normal or high workload),
- Events that triggered configuration changes (when a "low workload", or a "high workload" probability reaches a threshold).

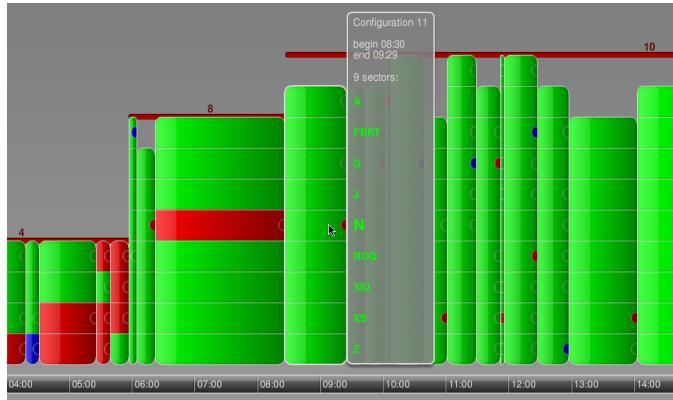


Figure 5: Overview of the sector configurations over one day.

Figure 5 represents a part of the global view. The sector background color represents the "global" workload so that durable overloads (red for danger) or underloads (blue) appear very clearly. The colored semi-circles represent the events that trigger a reconfiguration: red for high workload, blue for low workload. Red lines, constituting a limit for some configurations, express the constraints on these configurations in terms of working positions (staff limitations). The sector name appears only for the configuration flown over by the mouse pointer. When the user wants to know more about a specific sector, he can click on it or use a contextual menu to switch to the detailed view via a smooth transition using morphing and fish-eye unfolding, in order to keep the focus on the "selected" configuration and sector among many others.



Figure 6: Details of configurations over a given time span.

Figure 6 shows a part of the detailed view, where the configurations proposed by the model are displayed as stacks of sectors alongside a time scale, just as in the global view. A sector's global workload is represented by the color of its name. If it's overloaded, the whole sector is also emphasized in red. The colored bars at the right of some sectors symbolize the events. In the detailed view, configurations are folded by default (low level of information because there are a great many configurations), and may be unfolded when selected by the user (high level of information on a few chosen items) as shown in Figure 7. The awareness of connections between, and the evolution of, the sectors is reinforced by animated transitions between folded/unfolded states.

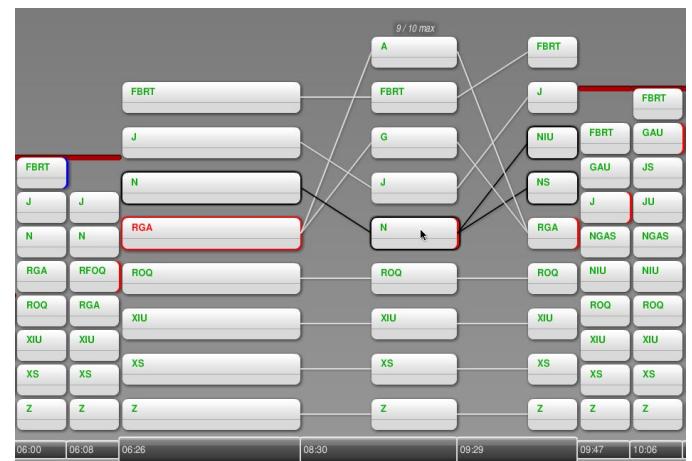


Figure 7: Details of reconfigurations over a given time span.

In Figure 7, the configuration at time $t = 08:30$ and the surrounding configurations are unfolded. The links between the sectors of these successive configurations are displayed in order to have a better idea of the inheritance.

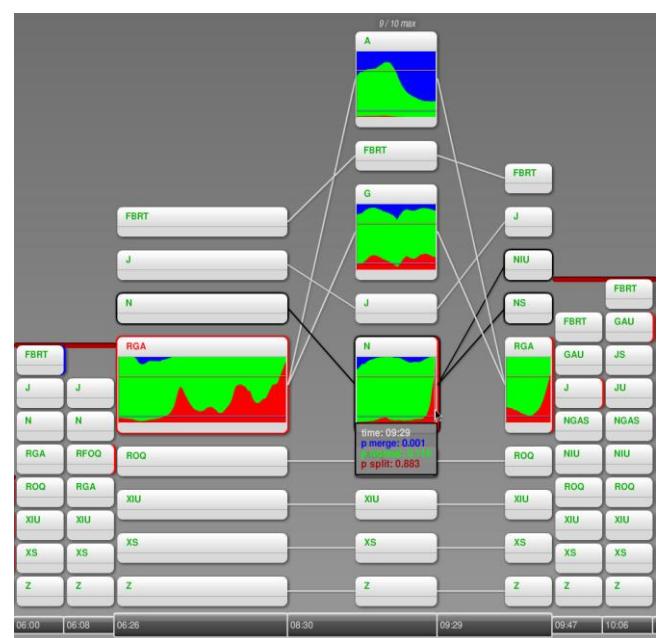


Figure 8: Details of reconfigurations with their associated metrics.

It is also possible to focus on specific control sectors: probability graphics are then displayed for the selected sectors (Figure 8), with a red area for the high workload probability, a green area for the normal workload probability, and a blue area for the low workload probability. The sum of the three probabilities being invariably 1, we stacked them on a single graph (the “high workload” probability is in the lowest area, and the “low workload” probability is in the upper area).

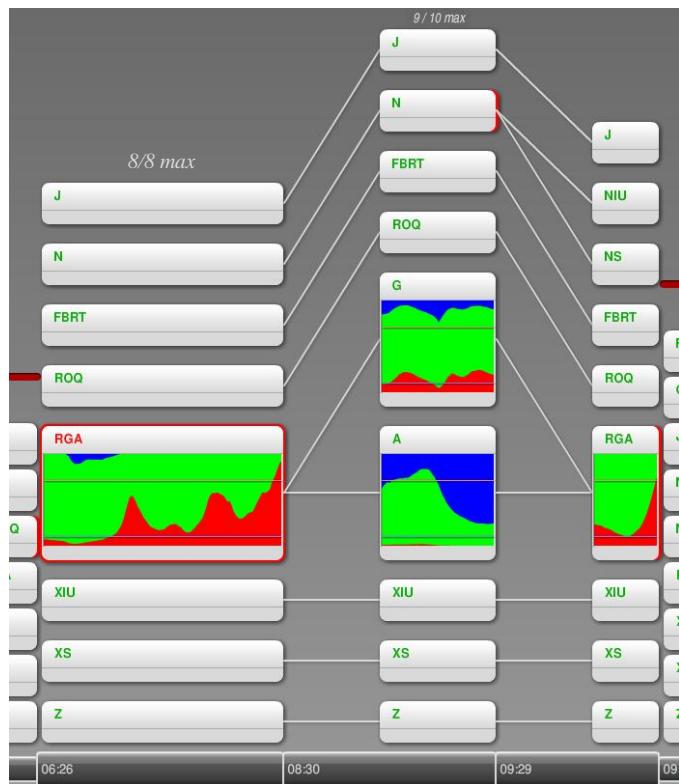


Figure 9: Untangled view with the overloaded RGA sector with no immediate solution.

The threshold values for the workload probabilities are also displayed (thin lines with the same color code as areas) so that it's easy to identify the sector that triggers the configuration changes. For example, considering the workload graph of the pointed sector N in Figure 8, one may see that the high workload probability reaches the decision threshold (probability 0.7) at time $t = 09:29$, triggering a reconfiguration where this overloaded sector is split into two smaller sectors NIU and NS. One may also see the sector RGA (in the configuration beginning at time $t = 06:26$) becoming overloaded a few minutes before the end of the configuration (08:30). Normally this situation should lead to an immediate reconfiguration. But at this moment, due to staff limitations, the number of working positions is limited to 8, so the algorithm cannot split this sector. It cannot find a better solution during the following few minutes, thus this sector is strongly emphasized in red to render this abnormal and durable situation clearly visible, whether folded (Figure 7) or unfolded (Figure 9). At 08:30, the staff limitation is raised to 10 working positions. Hence the algorithm can split the overloaded sector into two smaller sectors A and G.

The interactions with the interface are difficult to show here but their principle can be succinctly presented. They are divided into three categories: those which aim at improving the global presentation, those that help to keep the focus on the examined configurations while manipulating or selecting them among the many others and, finally, those that permit the acceleration of manipulations.

In the first category we can cite the ability to sort the sectors manually by dragging and dropping them inside their configuration. This interaction, requested by users at the beginning of the project, is practical when many sectors are opened and some of the links between sectors of two successive configurations may cross, which makes the view difficult to interpret (tangled lines in Figure 8 compared to untangled lines in Figure 9). The user can also choose to unfold three or five configurations (one or two before and after the selected configuration), which, again, was requested by the users (the ATCOs) during the evaluation of the prototype in order to improve, at a given moment, the visibility of the inheritance and return easily to a more compact view. The same goal led us to highlight the immediate inheritance of the sector flown over by the mouse pointer (see Figure 7, where the sectors N, NIU and NS have a dark drop shadow and their links to the pointed sector N are blackened).

In the second category, we can cite a zooming centered on the flown-over configuration and a centered unfolding mechanism divided into three steps: when a folded configuration is clicked, this configuration is first unfolded, and then the surrounding configurations are unfolded. During these two steps, the selected configuration is maintained at its original place. Finally, if necessary, there is an animated translation in order to ensure the visibility of the whole unfolded area.

In the third category, there are several choices of automated manipulations in a contextual menu, depending of the flown over component. Amongst them, there is the possibility to fold or unfold all the sectors of the flown over configuration or all the sectors directly linked to the flown over sector. These features were requested by ATCOs during the evaluation in order to have a rapid and accurate view of the workload in several sectors of a configuration or along an inheritance without having to click on each of them. It was also practical to be able to come back to an intermediary compact view (selected configurations are unfolded but their sectors are all folded as in Figure 7) without having to click on all the unfolded sectors or to fold all the configurations by clicking outside, and then unfold them again.

VII. CONCLUSION AND FUTURE WORK

In this paper, we have depicted the participatory design process that was used to create a Graphic Interface for new algorithms providing airspace configuration forecasts. Our approach was sequenced into several steps: study of the existing displays and ATC tasks in the current context, identification of the relevant information to be displayed in our new forecasting application, brainstorming work to sort out ideas with all the potential users, production of low fidelity (mainly paper) prototypes, assessment of these prototypes,

software implementation, iterative assessments and improvements of the software. We have presented the current version of the resulting research prototype that allows the users to visualize and interact with “off-line” data.

The participatory design approach was found efficient by the users group and other participants. Firstly, the users' feedback on the Graphic Interface and the capability of the algorithms to be operationally useful were highly positive. Secondly, since the users were involved throughout the design process, they also participated actively in the project and had a better understanding of the research issues that were addressed, or that still need to be explored.

Defining or identifying the operational concept (the final user's role and activities, interactions with other actors, etc) into which the proposed forecasting tool could fit, was not within the scope of this participatory design approach. However, some potential uses of the tool were evoked during the brainstorming sessions. Pre-tactical forecasts on a whole day of traffic, computed from Flight Plans, could be used in the current context by the FMP operator. On-line, real-time forecasts could fit in with the SESAR operational concept by allowing the complexity manager to mitigate the risk of overload in the ATC sectors. In that context, the proposed algorithms could allow the operator firstly to anticipate future overloads, and secondly, to check the incidence of alternative corrective measures on the workload and airspace configuration.

In the current version, the successive airspace configuration changes across the day were pre-computed for a whole day of traffic, using actual flight plans as input. In future works, we plan to run the computations on demand through the interface. Subsequent iterations will also address the dynamic re-calculation of the airspace configuration forecasts on a receding time horizon, as well as the introduction of some operational rules on the transition from one airspace configuration to the next (i.e. avoid too frequent recombinations of a same sector, transfer airspace modules from one ATC sector to another).

The continuation of this iterative approach was proposed as part of the DSNA contribution to the SESAR program, in work package 4.7.1 (“Complexity management in en-route”) and its industrial counter-part WP 10.8.1 (“Complexity assessment and resolution”), as well as in WP 7.5.4 (“Dynamic airspace

configuration”). Within this framework, it is expected that this approach involving potential users, HCI experts, and researchers, in the development of the research prototype will allow us to demonstrate, and validate/invalidate the workload model and partitioning algorithm that are proposed to provide more realistic airspace configuration forecasts.

REFERENCES

- [1] C. M. Bishop. Neural networks for pattern recognition. Oxford University Press, 1995.
- [2] D. Gianazza. Airspace configuration using air traffic complexity metrics. In 7th USA/Europe Seminar on Air Traffic Management Research and Development, 2007. best paper of "Dynamic Airspace Configuration" track.
- [3] D. Gianazza, C. Allignol, and N. Saporito. An efficient airspace configuration forecast. In Proceedings of the 8th USA/Europe Air Traffic Management R & D Seminar, 2009.
- [4] D. Gianazza. Forecasting workload and airspace configuration with neural networks and tree search methods. Submitted to Artificial Intelligence Journal.
- [5] D. Gianazza and K. Guittet. Evaluation of air traffic complexity metrics using neural networks and sector status. In Proceedings of the 2nd International Conference on Research in Air Transportation. ICRAT, 2006.
- [6] D. Gianazza and K. Guittet. Selection and evaluation of air traffic complexity metrics. In Proceedings of the 25th Digital Avionics Systems Conference. DASC, 2006.
- [7] D. Gianazza. Smoothed traffic complexity metrics for airspace configuration schedules. In Proceedings of the 3rd International Conference on Research in Air Transportation. ICRAT, 2008.
- [8] J. Greenbaum, M. Kyng, Eds. Design at Work: Cooperative Design of Computer Systems. L. Erlbaum Associates Inc 1992.
- [9] G.W. Furnas, Generalized fisheye views. SIGCHI Bull. 17, 4 (Apr. 1986), 16-23.
- [10] M. I. Jordan and C. Bishop. Neural Networks. CRC Press, 1997.
- [11] W. Mackay, A-L. Fayard, L. Frobert, L. Médini, (1998) Reinventing the Familiar: Exploring an Augmented Reality Design Space for Air Traffic Control. In Proceedings of ACM CHI '98 Human Factors in Computing Systems. Los Angeles, California: ACM/SIGCHI. ACM Press. pp 558-565.
- [12] B. D. Ripley. Pattern recognition and neural networks. Cambridge University Press, 1996.
- [13] SESAR project, the ATM Target Concept, SESAR definition phase – deliverable 3.
- [14] D. Schuler, A. Namioka, Participatory design: Principles and practices. Hillsdale, NJ: Erlbaum 1993.

Flight experience and executive functions predict flight simulator performance in general aviation pilots

Mickaël Causse^{1,2}, Frédéric Dehais¹

¹Centre Aéronautique et Spatial ISAE-SUPAERO ; Université de Toulouse, France

²Inserm ; Imagerie cérébrale et handicaps neurologiques UMR 825; F-31059 Toulouse, France

Université de Toulouse ; UPS ; Imagerie cérébrale et handicaps neurologiques UMR 825; CHU Purpan, Place du Dr Baylac, F-31059 Toulouse Cedex 9, France

Mickael.causse@isae.fr

Josette Pastor

Inserm ; Imagerie cérébrale et handicaps neurologiques

UMR 825; F-31059 Toulouse, France

Université de Toulouse ; UPS ; Imagerie cérébrale et handicaps neurologiques UMR 825; CHU Purpan, Place du Dr Baylac, F-31059 Toulouse Cedex 9, France

josette.pastor@inserm.fr

Abstract— Unlike professional pilots who are limited by the FAA's age rule, no age limit is defined in general aviation (GA). Some studies revealed significant aging issues on accident rates but these results are criticized. Our overall goal is to study how the effect of age on executive functions (EFs), high level cognitive abilities, impacts on the flying performance in GA pilots. This study relies on three components: EFs assessment, pilot characteristics (age, flight experience), and the navigation performance on a flight simulator. The results showed that contrary to age, reasoning, working memory (WM) and total flight experience were predictive of the flight performance. These results suggest that “cognitive age”, derived in this study by the cognitive evaluation, is a better mean than “chronological age” consideration to predict the ability to pilot, in particular because of the inter-individual variability of aging impact and the beneficial effect of the flight experience.

Keywords: piloting performance, executive functions, flight experience, decision making, normal aging.

I. INTRODUCTION

The population of GA pilots is getting older in the USA [1] and in European countries like France where forty one percent of private pilots are more than fifty (BEA¹, 2008). Unlike professional pilots who are limited by the FAA's age 65 rule, no such restriction exists for GA pilots. Moreover, contrary to commercial aviation (CA) pilots, GA pilots have not necessarily experienced a professional training, fly mostly on their own, without any co-pilot and very few assistance systems, have less support from the air traffic control and are more affected by weather conditions. Not surprisingly, in GA, the accident rate is considerably higher than in CA [2].

Several studies have revealed significant aging issues on accident rates in GA [3] [4] [5], though these results are called into question [6] [7]. The assessment of the cognitive functioning is a key issue in pilot's aging as long as its decline represents a much higher risk of accidents than sudden physical incapacitation [8]. A substantial literature focuses on the

evaluation of the cognitive state of pilots but its conclusions remain contradictory. Several reasons may explain the difficulty to draw a definitive conclusion on the effects of aging on flight performance in GA pilots. There is a great inter-individual variability in the deleterious effects of aging on cognition [9]; the evaluations performed in classical human factors studies are rather nonspecific in terms of explored cognitive functions and do not necessarily focus on the ones that are the most impacted by aging; very few researches attempt to link, in the same population, the cognitive performances to the flight abilities; the greatest part of the studies is interested on safety aspects like communications [10], or decision making during landing [11]; few researches are exclusively related to the GA population; finally, another source of complexity arises from the suspected compensative role on aging effects of the flight experience [12].

II. COGNITIVE FUNCTIONS AND PILOTING

Numerous studies have been conducted to attempt to link the cognitive functioning with the flight performance. Different measurements of cognitive efficiency have been identified as crucial to the piloting ability, for example: time-sharing [13], speed of processing [14], attention [15] or problem solving [16]. Cogscreen-AE [17] is among the most widely used cognitive tests batteries in pilots aging studies. It consists of a series of computerized cognitive tasks that evaluate a large set of cognitive functions. This battery has been shown to be able to successfully discriminate between neurologically impaired and cognitively intact pilots [18]. Some Cogscreen-AE variables were predictive of flight parameter violation in Russian CA pilots [19]. Furthermore, Taylor and colleagues [20] were able to predict 45% of the variance of the flight simulator performance with four Cogscreen-AE predictors (speed/WM, visual associative memory, motor coordination and tracking) in a cohort of 100 aviators (aged 50-69 years). Contrary to this latter study that involved Cogscreen-AE, a rather generalist battery in terms of explored cognitive functions, we propose to focus specifically on EFs. Indeed, these functions are the earliest ones to be impacted by aging [21] and represent excellent clues of aging effects on the

¹ French Accident Investigation Bureau.

cognitive performance. The study of EFs has appeared recently in aeronautics, for instance, Hardy [22] found significant age-related differences in pilots' executive functioning (*e.g.* inhibition, set-shifting) and Taylor [23] established a relationship between interference control and the ability to follow air traffic instructions.

III. EXECUTIVE FUNCTIONS AND PILOTING

EFs underlie goal-directed behavior and adaptation to novel and complex situations [24]. They allow the inhibition of automatic responses in favor of controlled and regulated behavior, in particular when automatic responses are no more adapted to the environment. Three major low level EFs are moderately correlated with one another, but clearly separable [25]: set-shifting between tasks or mental sets ("shifting"), inhibition of dominant or prepotent responses ("inhibition"), and updating and monitoring of WM representations ("updating"). The prefrontal cortex (PFC) plays a dominant role in the implementation of EFs that also encompass decision-making [26] or reasoning abilities [27]. According to our hypotheses, EFs are crucial to piloting. Indeed, this activity takes place in a dynamical and changing context where new information must be integrated and updated continuously. We assume that flying light aircraft with no autopilot and very few assistant systems (like the TCAS² or weather radar) presupposes a strong involvement of the EFs for handling the flight, to monitor the engine parameters, to plan the navigation, to maintain and update situation awareness and to correctly adapt to traffic and environmental changes and perform accurate decision-making by inhibiting wrong behavioral responses.

IV. EXECUTIVE FUNCTIONS AND NORMAL AGING

Functional neuroimaging brings evidence that the brain is subject to anatomical and physiological modifications in normal aging [28]. The prefrontal lobes appear to be the earliest cerebral regions to be affected [29] and may account for a great part to age-related cognitive changes [31]. Because the prefrontal lobes mainly implement EFs, aging is suspected to provoke a selective alteration of these latter, for example the reasoning [30], inhibition [31] or updating [32] abilities. However, the executive changes vary considerably across people. The complex interactions between the cerebral structures underlying EFs [9], sociocultural factors and genetic factors [33] may explain the heterogeneity of this decline.

In this experiment, we proposed to evaluate specifically the EFs, high level cognitive abilities that present a strong vulnerability to aging effects [21]. More precisely, we assessed three low levels EFs (shifting, inhibition and updating) and a more established general ability: the reasoning. The reasoning performance reflects fluid intelligence, that support processes relevant for many kinds of abilities (verbal, spatial, mathematical, problem solving *etc.*) and adaptation to novelty. It is a concept very close to the executive functioning [34] [35]. The speed of processing was also collected because it represents a reliable measure of general cognitive decline during aging. Finally, we have also taken into account age and

² Traffic Collision Avoidance System.

the total flight experience to assess their respective participation to the flight performance variation. Our hypothesis is that the "chronological age" is not a sufficient criterion to predict the piloting performance and that the "cognitive age", evaluated by the cognitive functioning, is a more relevant criterion.

V. METHODS

A. Participants

The participants were 24 private licensed pilots (mean age = 43.3 years, $SD = 13.6$) rated for visual flight conditions. The pilots that had no longer flown during the past two years were excluded because of the potential impact on flight simulator performance. Inclusion criteria were male, right handed, as evaluated by the Edinburgh handedness inventory [36], native French speakers, under or postgraduate. Non-inclusion criteria were expertise in logics, airline pilots and sensorial deficits, neurological, psychiatric or emotional disorders and/or being under the influence of any substance capable of affecting the central nervous system. All subjects received complete information on the study's goal and experimental conditions and gave their informed consent. Given that flight experience may moderate age related deficits in the performance of domain relevant task [12], we attempted to homogenize the flight experience distribution across the life span of our sample.

B. Flight performance

1) Navigation

The flight scenario has been setup in collaboration with flight instructors to reach a satisfying level of difficulty and realism. To familiarize the participants with the PC-based flight simulator and minimize learning effects in order to obtain reliable flight simulator performances, each volunteer underwent a training session. Before the navigation, they received the instructions, a flight plan and various technical information related to the aircraft (*e.g.* aircraft's crosswind limit). Basically, the scenario implied to take off, reached a waypoint with the help of the aircraft radio navigation system and finally, land on a given airport. The pilots were instructed that they were in charge of all the decisions and that they could only receive an informative weather report before landing. In order to increase the subject's workload, the pilots had to perform a mental arithmetic calculation of the ground speed (thanks to the embedded chronometer). Moreover, a failure of the compass was scheduled. After this failure, the pilots had to navigate thanks to the magnetic compass, which presents the particularity to be difficult to use as it is anti-directional. The flight scenario lasted approximately 45 min. The performance assessment was exclusively founded on the flight path deviations (FPD), expressed in terms of amount of angular deviation in the horizontal axis from the ideal flight path.

C. Neuropsychological battery

1) Target hitting

This test provides a basic psychomotor reaction time [37]. The instruction is to click as fast as possible on each target. The performance is measured by a velocity index inspired by the Fitts' law [38]. The index is the average ratio of the base 10

logarithm of the distance in pixels between two targets, divided by the time in seconds to go from the first target to the second.

2) The 2-back test.

The 2-back test aims at assessing working memory (WM), in particular maintenance and updating abilities [39]. Subjects view a continuous stream of stimuli and have to determine whether the current stimulus matches in a specific dimension (shape for our test) the stimulus 2-back in the sequence (Figure 1). For each condition, the percentage of correct responses was collected.



Figure 1. The 2-back test. The participant stated if the current shape match to the 2-back shape in the sequence thanks to the response box.

3) Deductive reasoning

The logical reasoning test has been used in a previous study to assess executive functioning [40]. The goal of the task is to solve syllogisms by choosing, among three suggested solutions, the one that allows concluding logically. Syllogisms are based on a logical argument in which one proposition (the conclusion) is inferred from a rule and another proposition (the premise). We used four existing forms of syllogisms: *modus ponendo ponens*, *modus tollendo tollens*, *setting the consequent to true and denying the antecedent*. Each participant had to solve 24 randomly displayed syllogisms. The measurement was the percentage of correct responses.

4) The computerized Wisconsin Card Sorting test

The Wisconsin Card Sorting test (WCST) [41] gives information on the subject's abstract reasoning, discrimination learning and shifting abilities [42]. The test version here was a computer implementation very similar to the clinical version of the WCST [43]. The participant must sort cards according to three different unknown categories (color, shape, number); an audio feedback indicated whether the response is correct or not (*yes/no*). When the participant categorized successfully ten cards, the target category was automatically changed. The task ended when six categories were achieved (color, shape, number, color, shape, number) or when the deck of 128 cards was used. The total numbers of perseverative errors (at least two unsuccessful sorting on the same category) was derived from the individual cards' records (Figure 2).

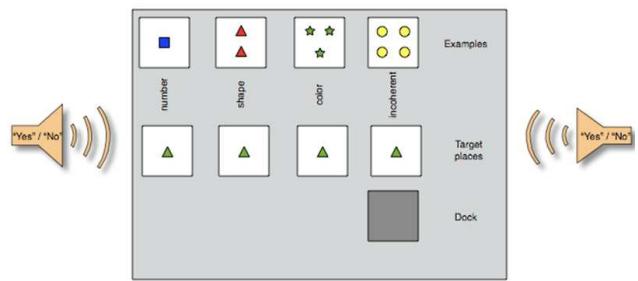


Figure 2. The Wisconsin card sorting test. The participant sorted the cards according to a specific dimension. An audio feedback informed if the sorting was correct or no.

5) Spatial stroop

Spatial Stroop tests generally assess the conflict between the meaning of a word naming a location (*e.g.* “below”) and the location where the word is displayed. The ability to restrain a response according to the localization of the word gives information on inhibition efficiency. This conflict appears to be provoked by the simultaneous activation of both motor cortices [44]. Our test encompasses four control conditions (Figure 3). “Stroop neutral meaning” (SNM): a motor answer is given with the appropriate hand according to the word meaning; “Stroop neutral position” (SNP): the response is given according to the location of a string of XXXXX, displayed at the left or the right of the screen; “Stroop meaning incompatible/compatible” (SMC/SMI): the response is given according to the meaning of the word, compatible or incompatible with its location at the screen. In order to get the pure effects of inhibition, the interference score was calculated to control reading and localization effects by:

$$SMI = \frac{SNP * SNM}{SNP + SNM}$$

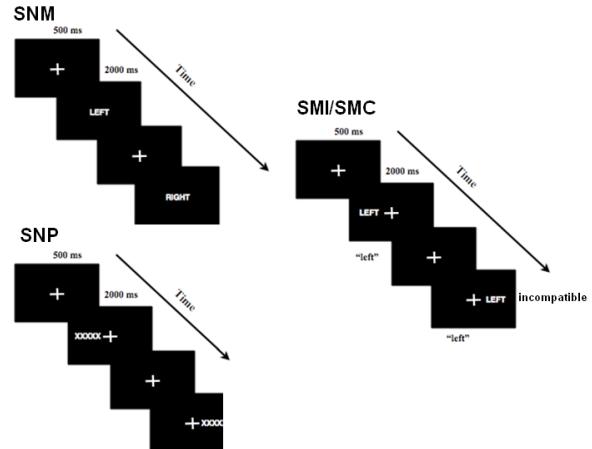


Figure 3. The four conditions of the spatial stroop. On the left: SNM, the participant pressed on the left/right button according to the meaning of the word; SNP, the participant pressed the left/right button of the response box according to the location at the screen of the pattern of XXXXX. On the right: SMC/SMI, the participant pressed the left/right button according to the meaning of the word, congruent or incongruent with its location at the screen.

D. Pilots characteristics

Age and total flight experience in hours were collected to assess their effects on the flight performance. We attempted to homogenize the flight experience distribution across the life span of our sample in order to minimize the perturbation of this parameter on the flight performance measurement.

VI. RESULTS

A. Statistical analysis

All data were analyzed with Statistica 7.1 (© StatSoft). The relationship between age and the total flight experience was examined thanks to Bravais-Pearson correlation. The ability of our control variables to predict the piloting performance was tested using exhaustive regression (ER) that searches for the best possible fit between a dependent variable and a set of potential explanatory variables. Contrary to classical stepwise approach, ER searches the entire space of potential models and returns those for which all parameter estimates are statistically significant. Thus, ER results are not affected by the order in which the variables are introduced in the model. The goodness of fit of the models was evaluated by the adjusted coefficient of determination r^2 .

B. Age and experience relationships

The mean total experience of our sample was of 1676 hours of flight (*Range* = 57-13000). The Bravais-Pearson correlation revealed that there was no relationship between age and total flight experience. However, in particular because of three aged pilots that owned a large total flight experience (respectively 61 and 13000 hours; 61 and 5000 hours; 58 and 6700 hours), the correlation was close to reach the significance ($p = .0561$, $r = +.39$).

C. Explanatory variables of the piloting performance

The mean FPD amplitude was 27.69 ($SD = 10.38$). The ER revealed that the performances of two cognitive abilities were predictive of the FPD: the reasoning and the WM (respectively, $p = .0083$, $F(1,15) = 9.20$, $p = .0395$, $F(1,15) = 5.08$. Moreover, the total flight experience was also a significant explanatory variable ($p = .0275$, $F(1,15) = 5.95$, see Figure 4).

The most the reasoning (see Figure 5 and Figure 6) and the WM abilities were efficient, the smaller was the FPD. In the same way, the most the pilots were experienced, the smaller was the FPD. The adjusted r^2 showed that this model explained 44.51% of the FPD.

As expected, the ER did not reveal any significant effect of age on the piloting performance ($p = .2488$, $F(1,15) = 5.95$). In the same way, the speed of processing and the two others

low level EFs, set-shifting and inhibition, were not predictive of the flight performance (respectively, $p = .5603$, $F(1,15) = 0.35$; $p = .8979$, $F(1,15) = 0.17$; $p = .9008$, $F(1,15) = 0.16$, see Figure 4).

It is interesting to note that the worst piloting performance (FPD = 52.01) has been done by a rather old pilot (62) with a very small total flight experience (90 hours) whereas two others aged pilots (both 61) with a high experience (13000 and 5000 hours) demonstrated correct flight performances (respectively FPD = 21.08 and 32.30).

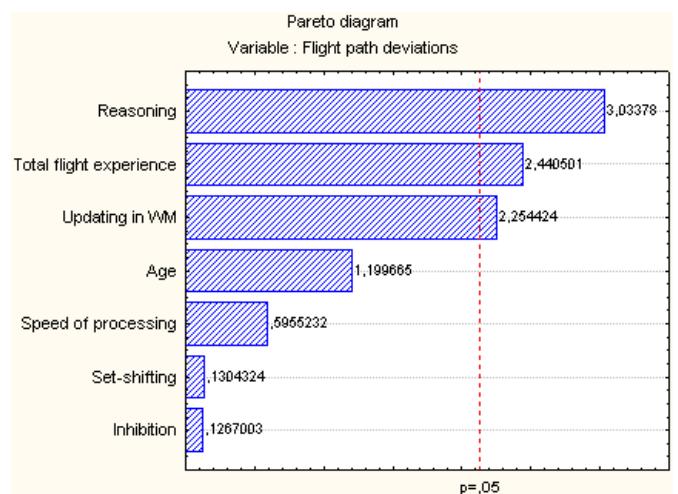


Figure 4. Synthesis of the ER. The Pareto diagram shows the three predictive variables of the flight performance: the reasoning abilities, the updating and the total flight performance.

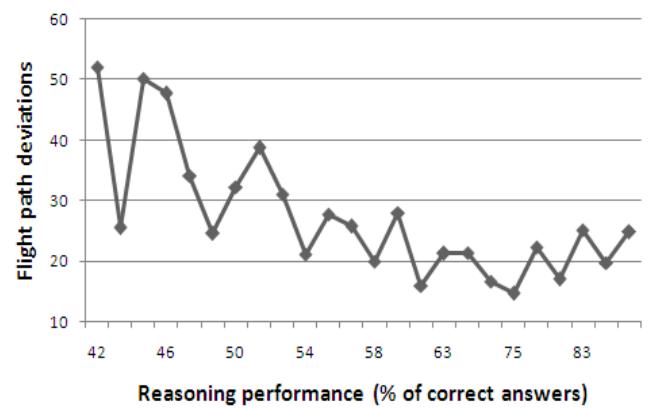


Figure 5. FPD as a function of the reasoning performances. The ER revealed that the reasoning performance predicts significantly the FPD.

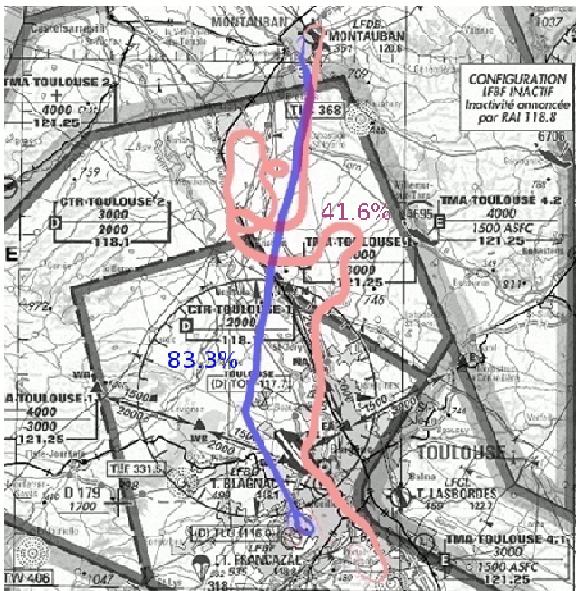


Figure 6. Flight path of two pilots and their respective reasoning performances. In blue, the pilot had a small flight path deviation and a good reasoning performance (83.3% of correct answers). In pink, the pilot had a large flight path deviation, he lost himself and flew by mistake above the Blagnac airport. His flight path deviation was very important and his reasoning performances were very low (41.6% of correct answers). Flight path are rendered with FromDady [45], the width of the line codes the altitude.

VII. DISCUSSION

A. Aging and piloting performance

According to our hypotheses and other authors [6] [7], the chronological age was not a relevant variable to predict the piloting performance. However, although the total flight experience was not correlated with age, it may have played a beneficial effect on some aged pilots. It is interesting to note that the worst piloting performance has been performed by a rather old pilot with a weak experience, whereas two others aged pilots, with a high experience, demonstrated quite good flight performances. In spite of these observations, our results raised the limitation of using the chronological age as a single criterion to decide if a given pilot is able to fly or not. In accordance with such statement, Schroeder [46] have pointed out the necessity to use neuropsychological tests rather than relying on chronological age.

B. Neuropsychological tests and piloting performance

The pilots performed a neuropsychological battery that taped three crucial low-level executive functions [25] plus reasoning and speed of processing. Finally, as revealed by the ER, reasoning performance was the variable the most predictive of the ability to pilot in our study. This result is not surprising, the reasoning abilities were strongly involved in our scenario. The pilots ought to perform numerous observations during the navigation to estimate their position and they had to use the radio navigation systems to reach a waypoint. Moreover, the scheduled compass failure required pilots to use the anti-directional magnetic compass as a backup. The utilization of this instrument is complex and could be a source

of difficulty. Although we did not assess precisely the errors associated with the use of this instrument, it seems likely that it has participated to increase the path deviation of some pilots. These results concerning the reasoning are in line with Wiggins and O'Hare [47] that have highlighted the links between reasoning performance, evaluated by a syllogism resolution (duncker's candle problem), and piloting performance. The reasoning performances reflect fluid reasoning, central cognitive ability linked with various types of mental activity (mental calculation, problem solving *etc.*) and essential to the adaptation to novel problems. Complex and novel problems cannot be solved directly by referring to a store of long-term knowledge but require analytic or fluid reasoning. The complexity of our scenario with unexpected event like the compass failure appears to have contributed to a strong involvement of reasoning abilities.

The total flight experience was also predictive of the FPD. In accordance with other studies [12], this data has confirmed the beneficial impact of experience on flight performance. This is coherent with Taylor's results [5] that showed that more expert pilots demonstrated better flight summary scores, especially in the communication and approach-to-landing. Moreover, this 3- year longitudinal study showed that aviation expertise was associated with less declines in flight simulator performance over time.

Finally, updating ability was also linked with the pilot's performances. This is coherent with our expectation. Indeed, the pilot's activity takes place in a dynamical and changing context where new information must be integrated and updated continuously. The updating performances are crucial in this context. Another study of Taylor et al. [20] found that the WM and the speed of processing were predictive of the piloting performance. We are partially in line with these results. We did not retrieve a significant effect of the speed of processing. The mean age of our sample was relatively low (43.3, $SD = 13.6$) and only seven participants of more than fifty were involved in the experiment. We may argue that more severe aging effects on speed of processing occur later in life, the sample of Taylor was more extreme and included participants from fifty to sixty-nine, these latter probably demonstrated more pronounced variations of speed of processing. Moreover, the task that we used to assess the speed of processing had a strong motor component that could have been less relevant to flight performance assessment.

Our overall results suggest that "cognitive age" is a better criterion than "chronological age" to predict the ability to fly and that reasoning and updating are good candidate to assess the cognitive age. The design of such neuropsychological batteries of tests that could be administrated during the pilot's periodic physical examinations could help to detect cognitive impairment associated with increased risk of accidents. Further research will include a larger sample of pilots and will be conducted on a more realistic flight simulator.

ACKNOWLEDGMENT

The authors wish to thank the Lasbordes Airfield ISAE staff. The study was supported by a "Gis Longévité" grant, by the DGA grant 0434019004707565 and the Midi-Pyrénées Regional Council grants 03012000 and 05006110.

REFERENCES

- [1] D. J. Hardy and R. Parasuraman (1997). Cognition and flight performance in older pilots. *Journal of experimental psychology applied*, 3, 313-348.
- [2] G. Li, and S. Baker (2007). Crash risk in general aviation. *Jama*, 297(14), 1596.
- [3] J. A. Y. Harkey (1996). Age-related changes in selected status variables in general aviation pilots. *Transportation Research Record: Journal of the Transportation Research Board*, 1517(-1), 37-43.
- [4] G. Kay (2001). Effects of aging on aviation-related cognitive function in commercial airline pilots. Paper presented at the In Proceedings of the 72nd Meeting of the Aerospace Medical Association, *Aviat Space Environ Med* 2001; 72 :241.
- [5] J. Taylor, Q. Kennedy, A. Noda, J. Yesavage (2007). Pilot age and expertise predict flight simulator performance: A 3-year longitudinal study. *Neurology*, 68(9), 648.
- [6] G. Li, S. Baker, J. Grabowski, Y. Qiang, M. McCarthy, G. Rebok (2003). Age, flight experience, and risk of crash involvement in a cohort of professional pilots. *American journal of epidemiology*, 157(10), 874-880.
- [7] G. Li, S. Baker, M. Lamb, J. Grabowski, G. Rebok (2002). Human factors in aviation crashes involving older pilots. *Aviation, space, and environmental medicine*, 73(2), 134-138.
- [8] D. Schroeder, H. Harris, D. Broach (2000). Pilot age and performance: an annotated bibliography. Oklahoma City, OK: Federal Aviation Administration, Civil Aeromedical Institute. Human Resources Research Division.
- [9] R. L. Buckner (2004). Memory and Executive Function in Aging and AD Multiple Factors that Cause Decline and Reserve Factors that Compensate. *Neuron*, 44(1), 195-208.
- [10] D. Morrow, W. Menard, H. Ridolfo, E. Stine-Morrow, T. Teller, D. Bryant (2003). Expertise, cognitive ability, and age effects on pilot communication. *The International Journal of Aviation Psychology*, 13(4), 345-371.
- [11] M. Wiggins and D. O'Hare (1995). Expertise in aeronautical weather-related decision making: A cross-sectional analysis of general aviation pilots. *Journal of Experimental Psychology: Applied*, 1(4), 305-320.
- [12] D. Morrow, W. Menard, E. Stine-Morrow, T. Teller, D. Bryant (2001). The influence of expertise and task factors on age differences in pilot communication. *Psychology and Aging*, 16(1), 31-46.
- [13] P. Tsang and T. Shaner (1998). Age, attention, expertise, and time-sharing performance. *Psychology and Aging*, 13, 323-347.
- [14] J. Taylor, J. Yesavage, D. Morrow, N. Dolhert, J. Brooks III, L. Poon (1994). The effects of information load and speech rate on younger and older aircraft pilots' ability to execute simulated air-traffic controller instructions. *The Journal of Gerontology*, 49(5), P191.
- [15] C. Knapp, R. Johnson (1996). F-16 Class A mishaps in the US Air Force, 1975-93. *Aviation, space, and environmental medicine*, 67(8), 777.
- [16] D. O'Hare, M. Wiggins, R. Batt, D. Morrison (1994). Cognitive failure analysis for aircraft accident investigation. *Ergonomics*, 37(11), 1855-1869.
- [17] R. Horst and G. Kay (1991). COGSCREEN- Personal computer-based tests of cognitive function for occupational medical certification.
- [18] G. Kay (1995). CogScreen: Aeromedical Edition. Odessa, Florida: Psychological Assessment Resources: Inc.
- [19] N. Yakimovitch, G. Strongin, V. Go'orushenko, D. Schroeder, G. Kay (1994). Flight performance and CogScreen test battery in Russian pilots.
- [20] J. Taylor, R. O'Hara, M. Mumenthaler, J. Yesavage (2000). Relationship of CogScreen-AE to flight simulator performance and pilot age. *Aviation, space, and environmental medicine*, 71(4), 373.
- [21] N. Raz (2000). Aging of the Brain and Its Impact on Cognitive Performance: Integration of Structural and Functional Findings. *The handbook of aging and cognition*.
- [22] D. Hardy, P. Satz, L.F. D'Elia, C.L. Uchiyama (2007). Age-Related Group and Individual Differences in Aircraft Pilot Cognition. *The International Journal of Aviation Psychology*, 17(1), 77-90.
- [23] J. Taylor, R. O'Hara, M. Mumenthaler, A. Rosen, J. Yesavage (2005). Cognitive ability, expertise, and age differences in following air-traffic control instructions. *Psychology and Aging*, 20(1), 117-133.
- [24] D.R. Royall, E.C Lauterbach, J.L. Cummings, A. Reeve, T.A Rummans, D.I. Kaufer et al. (2002). Executive Control Function A Review of Its Promise and Challenges for Clinical Research. A Report From the Committee on Research of the American Neuropsychiatric Association (Vol. 14, pp. 377-405): Am Neuropsych Assoc.
- [25] A. Miyake, N.P. Friedman, M.J. Emerson, A.H. Witzki, A. Howerter, T.D. Wager (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49-100.
- [26] A.G. Sanfey, R. Hastie, M.K. Colvin, J. Grafman (2003). Phineas gauged: decision-making and the human prefrontal cortex. *Neuropsychologia*, 41(9), 1218-1229.
- [27] R. Colom, V. Rubio, P. Shih, J. Santacreu (2006). Fluid intelligence, working memory and executive functioning. *PSICOTHEMA-OVIEDO-*, 18(4), 816.
- [28] R. Cabeza, N.D. Anderson, J.K. Locantore, A.R. McIntosh (2002). Aging gracefully: compensatory brain activity in high-performing older adults. *Neuroimage*, 17(3), 1394-1402.
- [29] D.J. Tisserand, J. Jolles (2003). Special issue on the involvement of prefrontal networks in cognitive ageing. *Cortex*, 39, 1107-1128.
- [30] W. De Neys, E. Van Gelder (2009). Logic and belief across the lifespan: the rise and fall of belief inhibition during syllogistic reasoning. *Developmental Science*, 12(1), 123-130.
- [31] R. West, G.C. Baylis (1998). Effects of increased response dominance and contextual disintegration on the Stroop interference effect in older adults. *Psychology and Aging*, 13, 206-217.
- [32] M. Van der Linden, S. Bredart, A. Beerten (1994). Age-related differences in updating working memory. *British journal of psychology*(1953), 85(1), 145-152.
- [33] I.E. Nagel, C. Chicherio, S.C. Li, T. von Oertzen, T. Sander, A. Villringer, et al. (2008). Human aging magnifies genetic effects on executive functioning and working memory. *Frontiers in Human Neuroscience*, 2.
- [34] S.L. Decker, S.K. Hill, R.S. Dean (2007). Evidence of construct similarity in executive functions and fluid reasoning abilities. *International Journal of Neuroscience*, 117(6), 735-748.
- [35] M. Roca, A. Parr, R. Thompson, A. Woolgar, T. Torralva, N. Antoun et al. (2009). Executive function and fluid intelligence after frontal lobe lesions. *Brain*.
- [36] R. Oldfield (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9(1), 97.
- [37] I. Loubinoux, D. Tombari, J. Pariente, A. Gerdelat-Mas, X. Franceries, E. Cassol et al. (2005). Modulation of behavior and cortical motor activity in healthy subjects by a chronic administration of a serotonin enhancer. *Neuroimage*, 27(2), 299-313.
- [38] P. Fitts (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47, 381-391.
- [39] Y. Chen, S. Mitra, F. Schlaghecken (2008). Sub-processes of working memory in the N-back task: an investigation using ERPs. *Clinical Neurophysiology*, 119(7), 1546-1559.
- [40] M. Causse, J-M. Sénard, J-F. Déméon, J. Pastor (2009). Monitoring cognitive and emotional processes through pupil and cardiac response during dynamic vs. logical task. *Applied Psychophysiology and Biofeedback*. DOI 10.1007/s10484-009-9115-0.
- [41] E.A. Berg (1948). A simple objective technique for measuring flexibility in thinking. *The Journal of general psychology*, 39, 15.
- [42] P. Eling, K. Derckx, R. Maes (2008). On the historical and conceptual background of the Wisconsin Card Sorting Test. *Brain and Cognition*.
- [43] R.K. Heaton (1981). A MANUAL for the Wisconsin Card Sorting Test. Odessa, Florida 33556: Psychological Assessment Resources, Inc.
- [44] M. DeSoto, M. Fabiani, D. Geary, G. Gratton (2001). When in doubt, do it both ways: Brain evidence of the simultaneous activation of conflicting motor responses in a spatial Stroop task. *Journal of Cognitive Neuroscience*, 13(4), 523-536.

- [45] C. Hurter, B. Tissoires, S. Conversy (2009). FromDaDy: spreading data across views to support iterative exploration of aircraft trajectories. *IEEE Transactions on Visualization and Computer Graphics*.
- [46] D. Schroeder, H. Harris, D. Broach (1999). Pilot age and performance: An annotated bibliography (1990-1999).
- [47] M. Wiggins, D. O'Hare (1995). Expertise in aeronautical weather-related decision making: A cross-sectional analysis of general aviation pilots. *Journal of Experimental Psychology: Applied*, 1(4), 305-320.

Economic Issues Provokes Hazardous Landing Decision-Making by Enhancing the Activity of "Emotional" Neural Pathways

Mickaël Causse, Frédéric Dehais, Josette Pastor
 Centre Aéronautique et Spatial ISAE-SUPAERO;
 Inserm; Imagerie cérébrale et handicaps neurologiques UMR
 825; F-31059 Toulouse, France
 Université de Toulouse; UPS; Imagerie cérébrale et handicaps
 neurologiques UMR 825; CHU Purpan, Place du Dr Baylac,
 F-31059 Toulouse Cedex 9, France
mickael.causse@isae.fr

Patrice Péran, Umberto Sabatini
 Laboratoire de Neuroimagerie, IRCCS Fondation Santa Lucia,
 Rome, 00149, Italie

Abstract— The analysis of aeronautical accidents highlights the fact that some airline pilots demonstrate a trend to land whereas the approach is not well stabilized. This behavior seems to be the consequence of various factors, including financial issues. Our hypothesis is that financial constraints modulate the brain circuitry of emotion and reward, in particular via the interactions between two prefrontal structures: the dorsolateral prefrontal cortex (DLPFC), main center of the executive functions (EFs), high level cognitive abilities, and the ventromedial prefrontal cortex (VMPFC), structure linked with the limbic system, major substratum of emotional processes. In our experiment, participants performed a simplified task of landing in which the level of uncertainty and the financial incentive were manipulated. A preliminary behavioral experiment ($n = 12$) was conducted. A similar second experiment using functional magnetic resonance imaging (fMRI) is in progress and a case study only is reported here. The behavioral data showed that the participants made more risky decision to land in the financial incentive condition in comparison to the neutral condition, where no financial incentive was delivered. This was particularly true when the uncertainty was high. The functional neuroimaging results showed that the reasoning performed in neutral condition resulted in enhanced activity in DLPFC. On the contrary, under the influence of the financial incentive, VMPFC activity was increased. These results showed the effectiveness of the financial incentive to bias decision-making toward a more risky and less rational behavior from a safety point of view. Functional neuroimaging data showed a shift from *cold* to *hot* reasoning in presence of the financial incentive, suggesting that pilot erroneous trend to land could be explained by a temporary perturbation of the decision-making process due to the negative emotional consequences associated with the go-around.

Keywords: decision making; emotion; reward; piloting performance.

Approach and landing are critical flight phases. They require formalised sequences of actions (e.g. to put and lock the gear down, to extend the flaps) and to follow an arrival procedure through several waypoints. They also require decision-making processes based upon rational elements like the maximum crosswind speed for a given aircraft. Uncertainty, a worsening factor since it generates psychological stress, can be high during landing. Moreover, several psychosocial factors lead some pilots to irrational decision-making, such as keeping landing on whereas all safety parameters are not respected [1]. According to the legislation, such hazardous conditions (e.g. thunderstorm, heavy rain, strong crosswind or windshear) require to go-around and to perform a new attempt to land more securely or a diversion to another airport. A study conducted by the MIT [2] demonstrates that in 2000 cases of approaches under thunderstorm conditions, two aircrews out of three keep on landing in spite of adverse meteorological conditions. This phenomenon, called *plan continuation error* [3], also exists in general aviation. Indeed, the BEA (the French Accident Investigation Bureau) revealed that this pilots' trend to land (the *get-home-it's syndrome*) have been responsible for more than 41.5 percent of casualties in light aircrafts [4].

II. ECONOMIC PRESSURES AND LANDING DECISION

Many experiments have addressed the difficulty for pilots to revise their flight plan and several cognitive and psychosocial explanatory hypotheses have been put forward [5] [6] [7] [8]. Another explanation to this trend to land in spite of adverse meteorological conditions or an unstabilized approach may resides in the impact of a large range of aversive consequences associated with the decision to go-around. Indeed, a go-around generates uncertainty and stress in the crew and the passengers, the pilot can feel it like a failure and it may lead to difficulties to reinsert the aircraft in the traffic. Moreover, a go-around has negative financial consequences for the company (fuel consumption in particular). An organisation's emphasis on productivity may unconsciously set

up goal conflicts with safety. The culture of the company weighs on safety: if it attaches a negative connotation to a go-around, it is an excellent candidate for landing accidents. One now-defunct airline used to pay passengers one dollar for each minute their flight was late until a crew attempted to land through a thunderstorm and crashed [9]. According to Orasanu [8], companies also emphasize fuel economy and getting passengers to their destinations rather diverting the flight, perhaps inadvertently sending mixed messages to their pilots concerning safety versus productivity. Those blurred messages create conflicting motives, which can affect unconsciously pilots' risk assessments and the course of action they choose. All these emotional pressures could alter the rational reasoning by shifting decision-making constraints from safety rules to economic optimization.

III. FROM COLD TO HOT DECISION MAKING

Neglected during the first half of the 20th century, the role of emotion on cognitive functioning has been recently fully established in the cognitive neurosciences. According to Koenig and Sander [10], this historical neglect of emotion is explained by the difficulty inherent to its investigation and by the influence of a scientifically-correct Cartesianism that considered the cognitive system as the "incarnation of reason". Today, it is commonly admitted that experiencing an emotion can trigger unconscious processes useful to decision-making, in particular when the uncertainty is high [11]. Many experiments put forward evidence of a strong interaction between the limbic system, "*emotional brain*", and other structures like the prefrontal cortex, the "*rational brain*". For instance, Drevet & Raichle [12] have shown the existence of a dynamic balance between regions of the limbic system (amygdala, posteromedial cortex, ventral anterior cingulate cortex) and regions more associated to EFs (dorsal anterior cingulate cortex, DLPFC). Similarly, Mayberg and colleagues [13] have put in evidence that an increased activity of limbic and paralimbic regions (subgenual cingulate, anterior insula) was proportional to the decrease in activity of neocortical regions (right prefrontal cortex, inferior parietal cortex) during the experience of sadness. These types of observations are supported by a study of Mitchell [14], which demonstrated that the activity of the DLPFC was inversely proportional to the VMPFC. A previous study of Goel & Dolan [15] has also highlighted this type of emotional and cognitive subdivision in the prefrontal cortex (PFC) in a reasoning task. In this study, when the reasoning task was performed without emotional induction, *cold reasoning*, DLPFC activations were found. When the same task was performed with emotional induction, *hot reasoning*, VMPFC activations were observed. All these studies allow to understand how emotion or stress are in relation to cognitive functions and how they can modulate the cognitive performance, in particular the EFs [16], mainly implemented within the PFC.

Our hypothesis is that *plan continuation error* may be, at least in part, related to a shift from *cold reasoning* to *hot reasoning*, in result of the different negative emotional consequences associated with the decision to go-around. *Cold reasoning* may be mainly supported by EFs whereas *hot reasoning* may be less rational from a safety point of view and more oriented toward company's financial interest. In a fully neuroergonomics approach, we propose to investigate this hypothesis by reproducing a simplified landing task performed in an fMRI.

IV. METHODS

A. Subjects

Two separate experiments were conducted. 12 physically and psychiatrically healthy volunteers were recruited from the local population to participate to the behavioral experiment (mean age = 28, $SD = 3.69$). 1 participant (age 28) was scanned in the fMRI. All subjects were right handed as measured by the Edinburgh handedness inventory [17]. The experiment was approved by the local ethic committee and an informed consent was obtained before participation.

B. From aircraft to fMRI

The task was based on a simplified reproduction of a real flight instrument, the ILS (Instrument Landing System). The participants were instructed that they were flying a plane during the landing phase and that like pilots, they were allowed to avoid landing if they believed that landing was unsafe (Figure 1). Decisions were based on two elements of the ILS: the localizer and the glide path, which provide lateral and vertical guidance to adjust the trajectory of the aircraft to the runway. This information was given by two rhombuses, like in real aircraft, displayed below and on the right of the PFD (Primary flight Display). It was explained to participants that the landing was safe when both rhombuses were close to the center of their axes, the farthest from the center the rhombuses were, the higher was the risk of crash. For each trial, the participants indicated their choice by pressing a button on the response pad. A first independent variable with two modalities was the degree of uncertainty, high or low, linked with the rhombuses position. The second independent variable was the type of incentive, neutral or financial. During the neutral condition, the incentive was only based on a feedback that gave information on the accuracy of the response. During the financial condition, a financial incentive was added to the one that gave feedback on the accuracy of the response. The task consisted of a set of 4 runs, 2 neutral, and 2 financial in which the level of uncertainty was manipulated according to the two modalities (high and low).

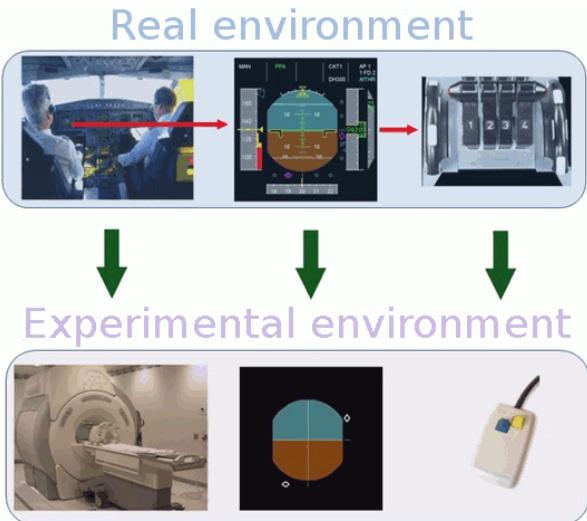


Figure 1. Simplified reproduction of the decision-making performed during the landing phase. In the upper part, the real environment. From left to right: the PFD within the cockpit, a zoom on the PFD and the throttle. In the bottom part, the experimental environment. From left to right: the fMRI, the simplified PFD (with only the two rhombuses of the ILS) and the response pad that allowed to accept to land or to perform a go-around.

C. Stimuli

1) The ILS

The stimuli was based on a 480x480 pixels simplified PFD with ILS and they reproduced a landing situation without external visibility. During the scan, they are displayed via back projection and an angled mirror in the head coil housing. Two different levels of uncertainty, depending of the positions of the two rhombuses, are randomly sorted within the 4 runs. In landings with low uncertainty, the decision-making was straightforward: either the rhombuses were very far from their respective center, requiring a go-around (likelihood of successful landing: 0%), either they were very close, requiring a landing acceptance (likelihood of successful landing: 100%). In landings with high uncertainty, rhombuses were borderline (not very far, not very close from the center) and the likelihoods (unknown by the subjects) of a successful landing or a crash is 50%. The positions of the two rhombuses were related to a score. Each axis was graduated with a 16 points scale, the most the rhombuses were far from the center, the higher was the score and the weaker was the likelihood to land securely (Figure 2).

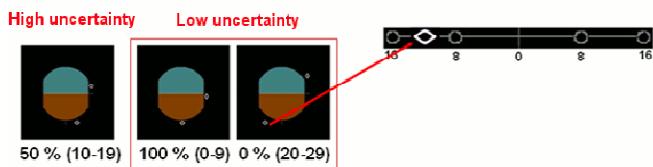


Figure 2. Categorization of the level of uncertainty according to the rhombuses positions. A score was computed according to the position of both rhombuses (the zoom gives information on only one rhombuse, the graduation was not displayed during the experimentation). The rhombuses position were counterbalanced to avoid laterality effects. The order of presentation of the stimuli was randomized.

2) Feedback

After each response, the participants received a feedback that informed on the response accuracy (OK, for a successful landing or a justified go-around; NO, for an erroneous decision to land or an unjustified go-around). At the end of each run, a global feedback indicated the percentage of correct responses (*safety score*). In addition, during the condition with financial incentive, a second feedback gave information about the financial consequences of each response (Figure 3). Moreover, at the end of this type of runs, another feedback indicated the cumulative amount of money won or loss (*financial score*).



Figure 3. The various feedbacks displayed after each decision making. Without incentive, only the accuracy feedback was delivered (OK/NO), with financial incentive, the monetary consequences are also displayed after the accuracy feedback.

D. The payoff matrix

During the financial incentive condition, negative emotional consequences associated with a go-around have been reproduced by a payoff matrix. This matrix was set up to bias responses in favor of landing acceptance. A go-around was systematically punished by a financial penalty. The penalty was less important (-2€) when the go-around was justified (in the case where rhombuses were very far from their center) than when it was unjustified (-5€). This systematic punishment of the decision to go-around reproduced the systematic negative consequences associated with this latter in real life. A successful landing was rewarded (+5€) whereas an erroneous decision to land was punished (-5€). The fact that the erroneous decision to go-around was less punished than the erroneous decision to land may appears counterintuitive but the matrix was set up in this way for two reasons. Firstly, in real life, the pilots know that crash and overrun are rather unlikely events whereas the negative consequences associated with a go-around are systematic. The analysis of unstabilized approach confirms that accidents are rather rare in spite of frequent risk taking [2]. Secondly, we could not reproduce the low frequency of accidents in an fMRI experiment because the cerebral signal associated with rare events could not emerge from a statistical point of view. For this reason, we were compelled to modulate the weight of the punishment rather than its frequency (Figure 4).

Case	GO	GO-A
Choice		
GO	+5€	-2€
GO-A	-5€	-2€

Figure 4. Payoff matrix biased in favor of landing acceptance. A successful landing pays 5€, an erroneous decision to land costs 2€, a justified go-around costs 2€ and an unjustified one costs 5€.

E. Experimental design for the behavioral study

We used a 2x2 factorial design crossing the financial incentive (neutral and financial) and the uncertainty (high, 50% chance of crash, or low, 100% or 0% chance of crash). Stimulus display and data acquisition were done with Cogent 2000 v125 running under Matlab environment (Matlab 7.2.0.232, R2006a, The MathWorks, USA). Two types of runs were presented during the experiment: neutral and financial ones. There were three likelihoods (0%, 100%, 50%) of successfully landing (40 trials for 0%; 40 trials for 100% and 80 trials for 50%), depending on both positions of the rhombuses displayed on the PFD. These likelihoods were unknown by the subjects. Each trial consisted in a presentation of the stimulus (3 s) during which the participant performed his decision thanks to a response pad, followed after a delay (5.5 s) by the feedback informing of the accuracy of the response (2 s). During the incentive condition the financial outcome was also displayed ({+5€}, {-5€} or {-2€}). Finally, an inter trial interval (2 s) was introduced. Response and financial feedback delivery was contingent upon the subject's responses

F. Experimental design for the fMRI study

The fMRI design was identical to that of the behavioral study excepted that the stimulus display duration was shorter (2.5 s) and that the delay duration (6-10 s) and the inter trial interval duration (3-9 s) were variables for neuroimaging technical issues. The long variable delay before the feedback allowed us to distinguish the hemodynamic signal associated with the decision taking during the stimulus presentation from the sustained signal associated with the reward uncertainty during the delay (Figure 5).

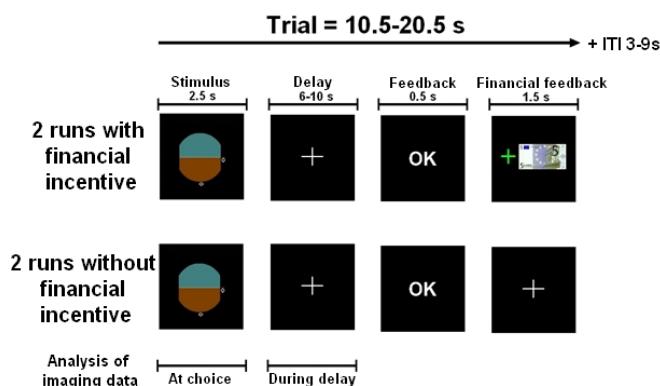


Figure 5. Experimental design. Four runs were performed by the participants (160 trials). The analysis of neuroimaging data was done during the stimulus presentation (at choice) and during the delay. The order of presentation of the run was fully counterbalanced to avoid order effects.

Before the experiment, participants performed two runs (neutral and financial) to become familiar with the task and the payoff matrix. The training is identical to the behavioral task.

G. fMRI image acquisition

The experiment was conducted at the Fondazione Santa Lucia (Rome). All the data were acquired in a single session on a 3 T Allegra scanner (Siemens Medical Solutions, Erlangen, Germany) with a maximum gradient strength of 40 mT/m, using a standard quadrature birdcage head coil for both RF transmission and RF reception. The fMRI data were acquired using a gradient echo-EPI, with 38 axial slices with a voxel size of $3 \times 3 \times 3.75$ mm³ (matrix size 64×64 ; FOV 192×192 mm²) in ascending order. The acquisition time was 2.47 s / 65 ms/ slice.

H. Analysis of fMRI data

Data analysis was performed within the Statistical Parametric Mapping analytic package (SPM5, Wellcome Department of Cognitive Neurology, London, UK). The data were sinc-interpolated in time. All images were re-aligned to the first acquired volume to correct head motion. Image was then spatially normalized and the transformation parameters were then applied to the functional volumes, smoothed with a (6*6*6*8 mm) isotropic Gaussian smoothing kernel. The preliminary analysis focused on non-specific effect of the financial incentive by collapsing reward regressor for the period of the stimulus and the delay and for every level of uncertainty. Thus two regressors were used: [low / high uncertainty, neutral], [low / high uncertainty, financial]

V. RESULTS

A. Statistical analysis

All behavioral data were analyzed with Statistica 7.1 (© StatSoft). The Kolmogorov-Smirnov goodness-of-fit test shown that data distribution was not normal, therefore, the effects of the financial incentive (neutral vs. financial), of the level of uncertainty (low vs. high) and their interactions on our dependant variables, the reaction times (RT) and the percentage of landing, were examined thanks to Friedman's ANOVA for overall effects and Sign test for paired analyses. The same type of analysis was also used to examine the effects of the type of stimulus (0%, 100% and 50%) on the same dependant variables.

B. Behavioral results

1) Effect of Uncertainty on reaction times

The Sign test revealed an effect of uncertainty on RT ($p < .023$). Higher uncertainty generated longer mean RT than low uncertainty (see Figure 6).

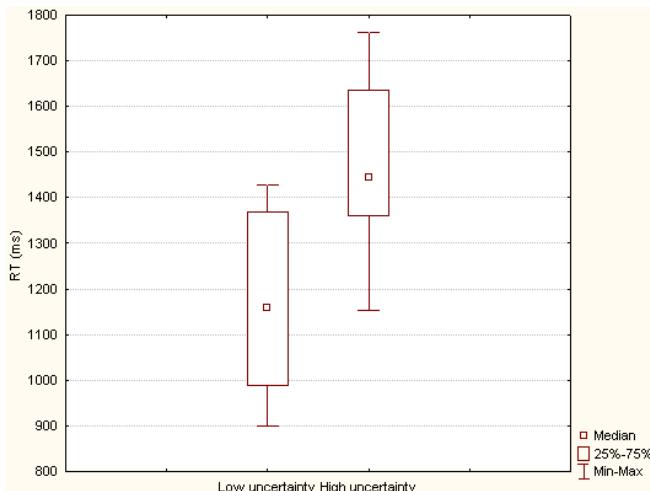


Figure 6. Reaction times (in ms) according to the two level of uncertainty (low and high).

2) Cross-effect of incentive and uncertainty on reaction times

The Friedman's ANOVA did not revealed an overall effect of the type of incentive on the RT. However, the Sign test revealed an effect of the financial incentive on RT ($p < .023$) for the stimuli where the landing was obviously possible (100% vs. 100%*). During the financial condition, the subjects RT were shorter than during the neutral condition, see Figure 7.

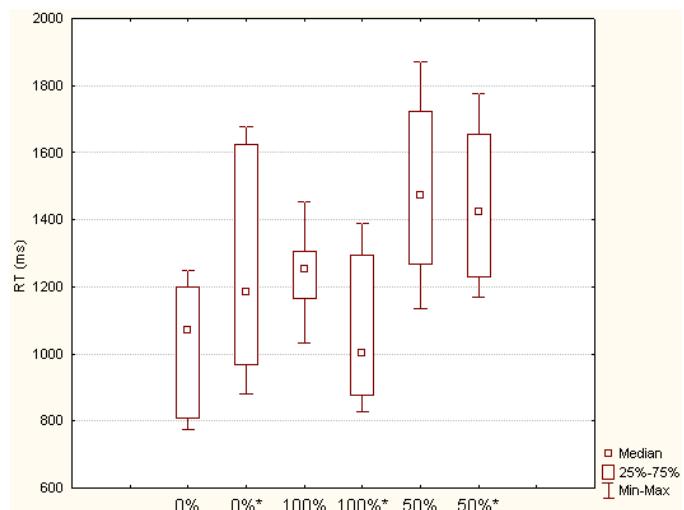


Figure 7. Reaction times (in ms) according to the 3 type of stimulus and for the two types of incentive. The asterisk indicates the presence of the financial incentive.

3) Cross-effect of incentive and uncertainty on decision making

In response to the asymmetric payoff matrix, subjects demonstrated a significant shift in the likelihood of accepting landings. More precisely, the Sign test showed that under uncertainty (50% vs. 50%*), the percentage of landing

acceptance increased ($p = .026$), from 32.09% ($SD = 12.06$) to 74.03% ($SD = 27.85$), see Figure 8.

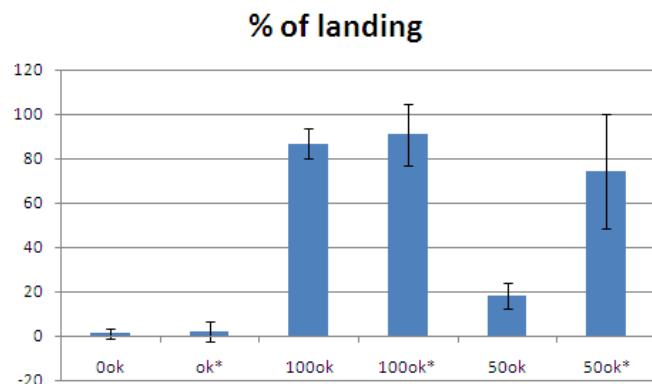


Figure 8. Percentage of landing acceptance according to the level of uncertainty and the type of financial incentive. Concerning the high uncertainty (50%), results showed a conservative behavior without incentive (under 50% of acceptance) and a risky behavior with incentive (beyond 50% of acceptance). The asterisk indicates the presence of the financial incentive.

C. Functional neuroimaging case study

The subject that performed the task in the scanner demonstrated behavioral results that were coherent with the behavioral group. The RT decreased in the financial incentive condition. Moreover the financial incentive generated a shift in the likelihood of accepting landings under high uncertainty, from 50% in the neutral condition to 85% of landing acceptance in the financial condition.

We investigated which brain regions were differently involved in decision-making under monetary incentive and in the neutral condition by performing overall contrasts that collapsed the time of choice and the time of the delay. The *cold reasoning*, performed during the neutral condition was associated with an increased activity in right DLPFC. On the contrary, the *hot reasoning*, performed under financial incentive was related with an increased activity in bilateral VMPFC (Figure 9).

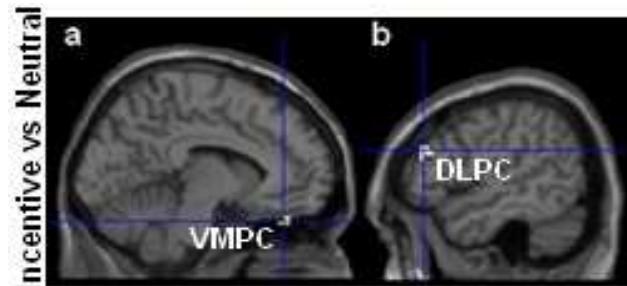


Figure 9. (A) Increased activation of the bilateral VMPFC (BA11) during stimulus and delay for monetary incentive vs. neutral. (B) Increased activation of the right DLPFC (BA46) during stimulus and delay for neutral vs. monetary incentive ($p < 0.01$; cluster > 15).

TABLE I. TALAIRACH COORDINATES, Z-VALUE AND CLUSTER SIZES (K) OF THE ACTIVATED ANATOMICAL STRUCTURES FOR THE CONTRASTS NEUTRAL VS. FINANCIAL AND FINANCIAL VS. NEUTRAL. ALL AREAS WERE SIGNIFICANT AT $P < .01$ UNCORRECTED.

Anatomical structures (Broadman's area)	Neutral - financial					Financial - neutral				
	Talairach coordinates				k	Talairach coordinates				k
Frontal	x	y	z	z-value		x	y	z	z-value	
DLPFC (BA46)	53	36	19	2.35	22	-12	42	-24	2.33	21
VMPFC (BA11)										

VI. CONCLUSION

We used an approach borrowed from neuroeconomics to investigate the impact of an economic pressure, namely the cost of a go-around, on landing decision-making. This preliminary work reports behavioral results and a case study in fMRI. The behavioral data confirmed the impact of the financial incentive. Firstly, subjects responded with a slightly faster response time for the financial incentive condition when the decision to land was obvious (100%), showing more precipitate responses. The decision to land in this context is rewarded by 5€ and this reduction of the RT may be interpreted as a search of the reward at the expense of a detailed analysis of the rhombuses positions. Secondly, the financial incentive has biased responses toward more risky decision-making. Whereas, under uncertainty, participants are rather conservative during the neutral condition (landing acceptance rate: 32.09%), they took more risky decisions under the influence of the biased payoff matrix (landing acceptance rate: 74.03%). The payoff matrix has associated the go-around with immediate negative consequences and participants became reluctant to do it.

The preliminary neuroimaging results confirmed that the change in decision-making entailed by the financial incentive was subserved by a shift from a cerebral region dedicated to reasoning (DLPFC) to a region involved in emotion processing (VMPFC). The behavioral data associated to these neuroimaging results are in favor of a shift from a *cold reasoning* under the neutral condition to a *hot reasoning* in presence of the financial incentive. According to us, this shift can be generalized to pilots and demonstrates that the erroneous trend to land whereas the approach is not stabilized is the result, at least for a part, of the different aversive negative consequences associated with the go-around decision, in particular the financial cost for the company. This is suggesting that this phenomenon may be explained by a temporary perturbation of decision-making process under an emotional factor. Data from fMRI sessions are currently analyzed and include a total of 16 participants.

REFERENCES

- [1] F. Dehais, C. Tessier, L. Christophe, and F. Reuzeau (2009). "The perseveration syndrome in the pilot's activity: guidelines and cognitive countermeasures" *HESSD* 2009.
- [2] D. Rhoda, and M Pawlak. (1999). "An assessment of thunderstorm penetrations and deviations by commercial aircraft in the terminal area". *Massachusetts Institute of Technology, Lincoln Laboratory, Project Report NASA/A-2, 3.*
- [3] J. Orasanu, L. Martin and J. Davison (1998). "Errors in Aviation Decision Making: Bad Decisions or Bad Luck?" *RECON(20020063485).*
- [4] BEA. (2000). "Objectif : destination". Technical report.
- [5] J. Goh and D. Wiegmann (2002). "Human factors analysis of accidents involving visual flight rules flight into adverse weather". *Aviation, space, and environmental medicine*, 73(8), 817.
- [6] E. Muthard and D. Wickens (2002). "Change detection after preliminary flight decisions: Linking planning errors to biases in plan monitoring". *Human Factors and Ergonomics Society Annual Meeting Proceedings*.
- [7] D. O'Hare and T. Smitheram (1995). "Pressing on" into deteriorating conditions: An application of behavioral decision theory to pilot decision making". *The International Journal of Aviation Psychology*, 5(4), 351-370.
- [8] J. Orasanu, N. Ames, L. Martin and J. Davison (2001). "Factors in Aviation Accidents: Decision Errors". *Linking expertise and naturalistic decision making*, 209.
- [9] J. Nance (1986). "Blind trust: How deregulation has jeopardized airline safety and what you can do about it". *W. Morrow and Co.*, New York.
- [10] D. Sander and O. Koenig (2002). "No inferiority complex in the study of emotion complexity: A cognitive neuroscience computational architecture of emotion". *Cognitive Science Quarterly*, 2, 249-272.
- [11] A. Damasio (1995). "L'erreur de Descartes". *Éditions Odile Jacob*.
- [12] W. Drevets and M. Raichle (1998). "Reciprocal suppression of regional blood flow during emotional versus higher cognitive processes: Implications for interactions between emotion and cognition". *Cognition and Emotion*, 12(3), 353-385.
- [13] H. Mayberg, M. Liotti, S. Brannan, S. McGinnis, R. Mahurin, P. Jerabek et al. (1999). "Reciprocal limbic-cortical function and negative mood: converging PET findings in depression and normal sadness". *American Journal of Psychiatry*, 156(5), 675.
- [14] R. Mitchell and L. Phillips (2007). "The psychological, neurochemical and functional neuroanatomical mediators of the effects of positive and negative mood on executive functions". *Neuropsychologia*, 45(4), 617-629.
- [15] V. Goel and R. Dolan (2003). "Reciprocal neural response within lateral and ventral medial prefrontal cortex during hot and cold reasoning". *Neuroimage*, 20(4), 2314-2321.
- [16] D. Schoofs, O. Wolf and T. Smeets (2009). "Cold pressor stress impairs performance on working memory tasks requiring executive functions in healthy young men". *Behavioral Neuroscience*, 123(5), 1066-1075.
- [17] R. Oldfield (1971). "The assessment and analysis of handedness: the Edinburgh inventory". *Neuropsychologia*, 9(1), 97.

Empirical Analysis of Air Traffic Controller Dynamics

Yanjun Wang^{†*‡}, Frizo Vormer^{*}, Minghua Hu[†], Vu Duong^{*}

[†]Nanjing University of Aeronautics and Astronautics, Nanjing, China

^{*}EUROCONTROL, Bretigny-sur-Orge, France

[‡]Telecom-ParisTech, Paris, France

{yanjun.wang.ext, frizo.vormer, vu.duong}@eurocontrol.int

minghuahu@263.net

Abstract—This paper addresses an empirical analysis of Air Traffic Controller activities using a human dynamics and complex systems approach. Workload metrics have been long well investigated from a cognitive engineering, human factors approach, and have been widely used as an indicator of controller's activity levels. However, the dynamical property of workload is still unknown, which make it difficult to predict workload ahead of time. Recent investigations on human dynamics show several empirical evidences that, different from common belief respecting random-based Poisson distributions, patterns of human activities fit into power law distribution with heavy tail patterns. Our hypothesis lies upon the question whether or not controller's dynamics obeys the same power law pattern. Our first attempt consists in analyzing the temporal characteristics of controller activities, in term of communication activities. The analysis on ATCOSIM Air Traffic Control Simulation Speech corpus shows that inter-communication times do follow a heavy tail pattern. Over certain thresholds, the distribution of inter-communication times approximates power-law decaying, and the correlations between communication events and traffic activities are influenced by the time-scale selected. However, the meanings of the thresholds are not interpretable due to the lack of available information.

Keywords- air traffic control; human dynamics; complex systems;

I. INTRODUCTION

Despite the wider and wider range of automation that has been introduced into Air Traffic Management (ATM) systems, scenarios in both SESAR and NextGen concepts still reckon that air traffic controllers (ATCO) continue to constitute the core function of the future system. As the decision-maker and executor of the system, the performance of the controller is closely interconnected with the system safety and efficiency. The prediction of ATCO performance with respect to traffic activities is therefore of quantifiable importance.

It has been well known that workload is one of the main factors affecting controller's performance, and some research efforts have been focusing on measuring and predicting ATCO workload. Earliest work was based on queuing theory and examination of controller routine work. A queuing

model was proposed by Schmidt et al. based on the hypothesis of the single-channel of man's information-processing activity, trying to quantify and predict the workload factors affecting controller performance [1-3]. The prevalent approach to measure workload is based on the controllers' subjective rating [4]. Controller are asked to report the workload rate that they were experiencing either they are controlling traffic or just afterwards. On-line ratings could distract the controller from perceiving and controlling traffic and could influence the workload results. Whereas for the workload obtained after work, it may fail to capture the essential property of workload as it emerges from the complex interaction of current traffic situation and controller. In the workload modeling and predicting front, useful information can be found in Loft et al. [5], in which the authors pointed out the difficulties and shortcomings of existing studies and concluded with several suggestions, leading us to the study of the dynamics properties of workload incorporated with controller strategies management.

While the progress on the ATC workload analysis has been consequent, much of the available work show difficulties in the predictability of workload level. Given the fact that workload is one of the factors affecting human activities, from a system perspective, it is the human actions that influence the system operation. *The activity is a coherent system of internal processes and external behavior and motivation that are combined directed to achieve conscious goals*[6]. In the ATM system, voice communication was the primary means used by controller to control air traffic before the emerging of digital data communication between controller and aircraft, such as Controller Pilot Data Link Communication (CPDLC). However, it is still the only channel for information flow between pilots and controllers in most control centers. Under such circumstance, controller voice communications activities have direct impact on the whole system evolution. A set of communication activities can be seen as control strategies, which are the result of controller mental and physical efforts after the assessment of current situation according to their experiences.

Analysis of communication data has a long history. In the past, communication events were extensively used to measure workload [4, 7-10]. In [4], Manning et al. have

examined the relationship between communication events, subjective workload and objective task-load measures. The communication events used in their study were total number of communications, total time spent communicating, average time spent for an individual communication, and communication content. Although some measures of communication events are highly correlated with workload, the analysis does not make a unique contribution to the workload evaluation [4, 11]. It should be noticed that the focus of the above mentioned work is the relationship between communication events and workload, rather than the dynamics of communication activities. In this paper, we are interested in the patterns of communications of controllers; in particular the temporal behaviors of the depicted activities under the assumption that voice communications do reflect the activity level of controllers. We are also interested in finding a relationship between controllers' communication activities and traffic activities, knowing that traffic complexity is not necessarily the sole factor that drives ATCO communication events but emergencies and non-nominal events, which are randomly distributed.

The analysis is performed on the ATCOSIM Air Traffic Control Simulation Speech corpus of EUROCONTROL. The particular quantity we focus on is the inter-communication times, defined by the time differences between two consecutive speeches of the same controller. First, a literature review on human dynamics and how it relates to air traffic control are given in Section II. Then, we give a general description of data used in this study, and examine the relationship between traffic activities and communication activities in Section III. The statistical results are presented in Section IV. Finally, we list several questions and suggestions Section V, basically why our initial results demonstrate that controller dynamics could not be identified by the only use of current communication data.

II. RELATED WORK AND HYPOTHESES

In fact, the research on human behavior can be traced back to early 20th century[12]. Human activity as an academic subject has been extensively studied in psychology, sociology and physiological psychology and among the others. The difficulty of collecting experimental and real data had limited the quantitative investigation of human activity, which resulted in that the hypotheses and conclusions were given in qualitative. Due to the rapid development in electronic information technology, human activities data can be easily record which provides a great platform for studying human behavior. Until recently, the temporal characteristic of human actions had been thought to be randomly distributed. The basic assumption of human dynamics models, used from communications to risk assessment, had been that the temporal characteristics of human activities could be approximated by Poisson processes. However, there is increasing evidences showing that the inter-event times, defined by the time difference between two consecutive activities, indeed follow non-Poisson statistical distribution. It can be well characterized by heavy-tailed patterns, with bursts of rapidly occurring events separated by long periods

of inactivity. The analysis resulted from large empirical data sets, including human correspondence [13], email communication [14], human printing behavior [15], and online films rating [16], demonstrate that the distribution of inter-activities times can be well approximated by a power-law form with different exponents (See Figure 1 for example). For the first time, a priority queuing model was built by Barabasi to show the bursty nature of human activity rooted from the decision-based queuing process when human execute tasks [17]. Malmgren et al argued that the correspondence patterns are better described by a lognormal distribution rather than a power-law distribution [18]. They also constructed a double-chain Markov model for formulating the cascading non-homogeneous Poisson process, demonstrating that the human correspondence patterns are well described by the circadian cycle, task repetition and changing communication needs [14, 18, 19].

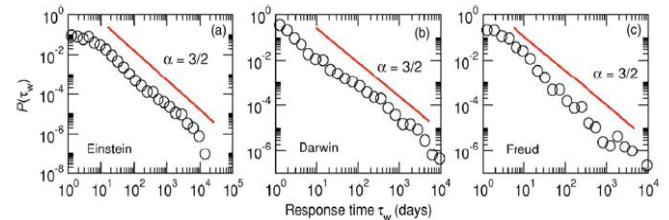


Figure 1. The distributions of respond times for the letters replied by Einstein, Darwin and Freud respectively. It is an exact copy from[20].

Analysis from mobile phone data sets demonstrates that the human trajectories show a high degree of temporal and spatial regularity [21]. In[22] the limits of predictability in human mobility has been studied by measuring the entropy of human trajectory. It was found that there is a 93% potential predictability in human mobility across the whole data. The underlying similarity among human actions indicates that there exists the same law, which governs human activity.

Although the studies in human dynamics have been impressive, one may argue that it is necessary to consider the following specific features of the air traffic controller's work when compared with the previously mentioned activities.

1. Dependence on environmental conditions: The main goal of ATCOs is to ensure the aircraft under jurisdiction reach their destinations respectively while adhere to the separation standards and operation regulations. Then, the characteristics of sector configuration, operational procedures, and air traffic are the main objective factors that may determine controller's behavior. Hence, the activities dynamics should be sector-specific and thus depending on the sector configuration, procedures, and traffic.

2. Urgency or time pressure: The air traffic controller has to complete many tasks to meet the rapidly changed situation. Compared with daily activities, such as email communication, many of controller's tasks are more time-pressuring. The competent controller has the ability to appropriately utilize the resources in the finite time. It is the strategies which the controller uses to maintain acceptable workload and performance level that determine his/her activities.

In spite of the above listed issues in air traffic control, controllers still have their flexibility to manage the resources, including airspace/airport resources and their own resources. There still exist common points among the controllers. For example, Histon et al. [23, 24] showed that a recognized underlying structure could act as the basis for abstractions internal to the controller, which can simplify the controller's working mental model. Standard flow, critical points, grouping, and responsibility are the four common type structure-based abstractions. The reduction of the "order" is the most effective way to mitigate cognitive complexity. However, the quantified description of the mechanisms by which controller uses to manage the air traffic are still unknown. From the human dynamics perspective, we propose the following hypotheses for the study in air traffic controller dynamics.

I. There exist similar activities patterns among air traffic controllers. The temporal characteristics of communication activities will be examined in this paper.

II. Network dynamics approach is appropriate to investigate the air traffic controller dynamics. This hypothesis is to be studied in future analyses.

III. METHOD

A. ATCOSIM Dataset

The controller communication data used in this study is from the ATCOSIM Air Traffic Control Simulation Speech corpus of EUROCONTROL Experimental Centre. The aim of the ATCOSIM is to provide a speech database of non-

prompted and clean ATC operator speech. It consists of ten hours communication data, which were recorded during ATC real-time simulations [25, 26]. These simulations were conducted between 20/01/1997 and 14/02/1997, with the aim to evaluate the concept of RVSM (reduced Vertical Separation Minimum) in Europe. For the purpose of ATCOSIM, only controllers' voice was recorded and analyzed. Considering the traffic initialization phase with little speech, the first half-hour of traffic was not recorded. Hence, each record consists of *circa* one hour of communication data. Both speech signal data and transcription of the utterance, together with the complete annotation and meta-data for all utterances, can be found in the database. ATCOSIM dataset also gives information about communication activities, including the speech start time, duration of the speech, and the content of the speech.

Six simulation exercises, which were conducted by four controllers, are considered in our study. The general information of these exercises is given in Table I.

TABLE I. GENERAL DESCRIPTION OF SIMULATION EXERCISES USED

Exercise ID	ATCO ID	Recording Time	Number of A/C under Control	Number of communication events
zf1_07	zf1	58'15"	66	211
zf2_07	zf2	64'30"	65	222
sm1_07	sm1	57'40"	66	165
sm2_07	sm2	59'16"	66	235
zf1_08	zf1	47'41"	61	196
sm2_07	sm2	56'34"	62	215

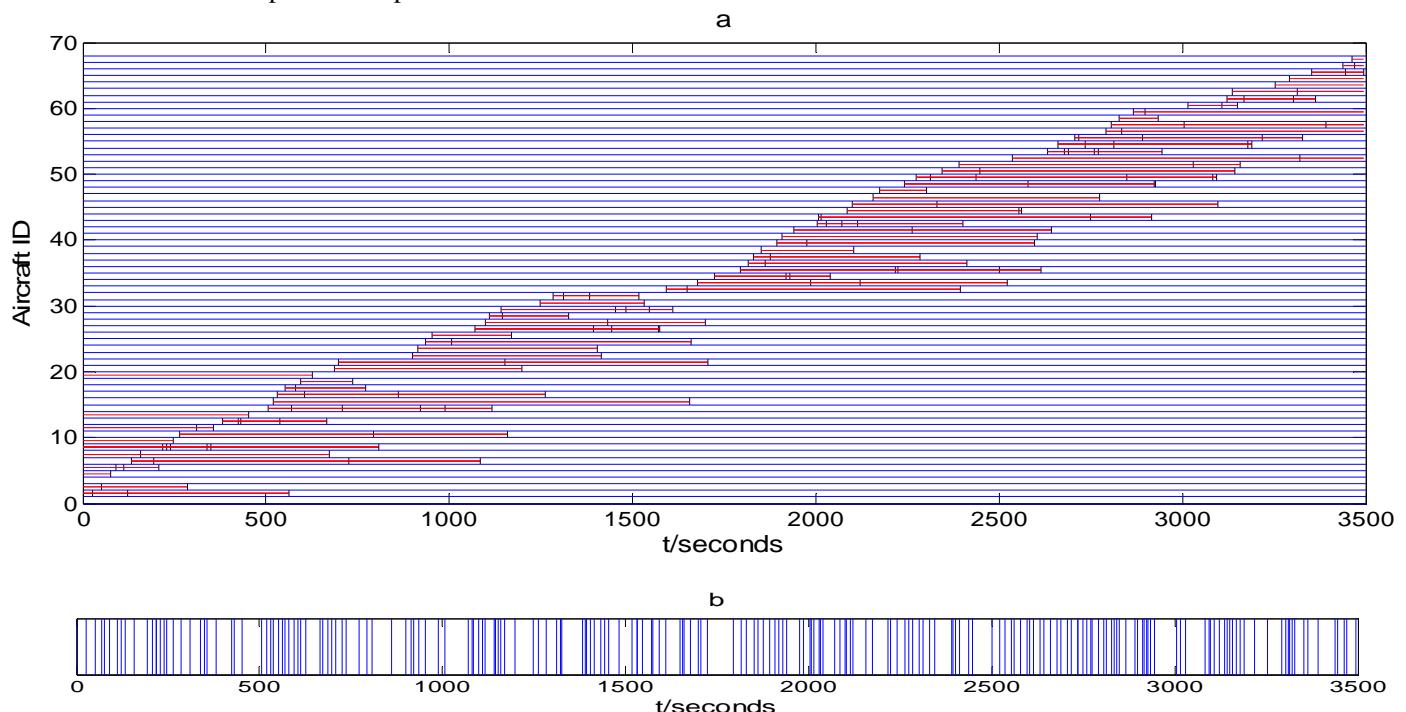


Figure 2. Historical communication activities constructed from controller speeches. a, Display of communication events according to the aircraft. Each horizontal red line stands for a different aircraft, with each vertical black line corresponding to a communication event. The length of red line denotes the duration that aircraft stays in the section during the exercise. b, The succession of communication activities of controller, with each blue vertical line represents a communication event over time.

B. Traffic activities and communication activities

The ATCOSIM dataset does not provide any information about airspace configuration or traffic scenarios, portraying the pictures of traffic situations from a voice communication dataset is not straightforward. The only way to acquire the traffic information is to perform deep analysis from controller's speech data by identifying aircraft call-signs spelled out by ATCO's during standard transfer instructions (when aircraft is entering or leaving the sector). In other words, the duration of aircraft flying through the sector can be deduced by means of timing the differences between transfer out and transfer in instructions. However, *circa* 5.2% of controller messages have not included aircraft call signs. Fig. 2 shows an example of communication events in exercise zf1_07.

The communication activity is defined as the event that controller press the push-to-talk button and hold in order to send the transmissions to aircraft, disregarding the contents of the transmissions. There are several empty transmissions in the database, which are also seen as the complete communication activities.

IV. RESULTS

A. Correlations between traffic activities and communication activities

Given the transfer in/out times of each aircraft, the number of aircraft under control within predefined duration is obtained by iteration. The dynamics on traffic volume and communication events are illustrated as in Fig. 3.

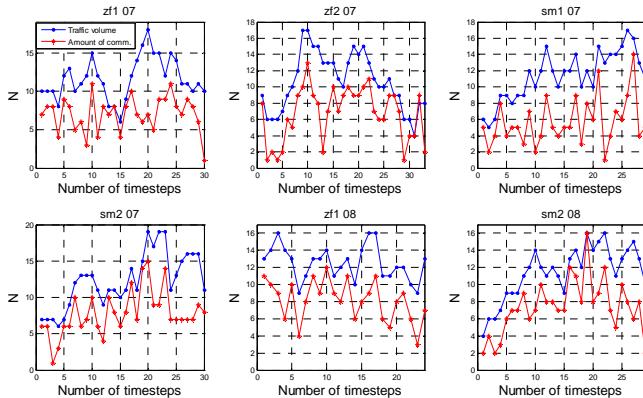


Figure 3. Statistical results on traffic volume and communication activities (time step: 2 minutes). The blue ● markers stand for the number of aircraft in the sector, while the red * denote the number of communication events in each time step.

It can be seen in Fig. 3 that the communication events vary with the change of number of aircraft in the sector. With the obtained number of aircraft and the number of communication events, we calculated the correlations between them. Table II shows these correlations with different time steps. It is found that although the total number of aircraft under control is highly correlated with the number of communication events, which is in agreement

with previous study [10], the strength of the relationships change with the time step.

TABLE II. CORRELATIONS BETWEEN TRAFFIC AMOUNT AND COMMUNICATION EVENTS. (WITH DIFFERENT Timestep)

	zf1_07	zf2_07	sm1_07	sm2_07	zf1_08	sm2_08
1 min	0.3207	0.6064*	0.4293*	0.5456*	0.5398*	0.5859*
2 min	0.3033	0.7520*	0.5849*	0.7162*	0.6840*	0.7586*
3 min	0.4863	0.8153*	0.6488	0.7724*	0.6344	0.8311*
5 min	0.3638	0.8836*	0.7340	0.8872	0.7676	0.8212*
10 min	0.7644	0.9531*	0.7183	0.8834*	0.6601	0.9481

* Correlation is significant at $p < 0.001$ level

Then, we analyze the number of aircraft as a function of the number of communication events. The aircraft without transfer in or transfer out instructions are discarded. As Fig. 4 shows, a large number of aircraft received less than 4 messages, while a small fraction of aircraft tends to get more. On average, less than 20% aircraft are communicated over 4 times (31% for sm2_07). After this initial drop however, few aircraft received 7 messages. Aircraft with more than 8 messages are supposed to have been in abnormal situations. The differences of maximum number of aircraft between exercises could be interpreted with the different types of sectors. For example, the sectors in the exercises sm1_07 may be the en-route sector, whereas the sector in sm2_07 is an approach sectors with few flights flying-over. We note that the discreetness and sparseness of the data does not allow us to prove the number of aircraft decay follows a power-law, but the shapes of curves do suggest this. Hence, the graphical analysis of Fig 4 will be considered for further investigations.

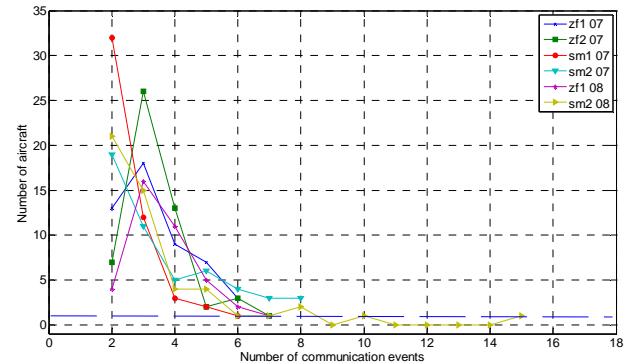


Figure 4. Number of aircraft as a function of the number of messages received.

To explain the reasons for that few aircraft received more than four messages, we need both traffic situation information and communication contents. The former will be used to investigate the potential conflicts, whereas the latter is for the deduction of air traffic controller's strategies for conflict resolution. However, it is difficult to reconstruct the traffic situation by using of the available database. We analyze the content of each transmission which is received

by the aircraft with more than four messages. Besides of the normal hand in/out communications, the rest of the transmissions are categorized into five groups according to its content.

- Inquiry: To ask or confirm the current flight level, speed, destination, etc.
- Altitude Change: Transmissions about changing flight levels
- Heading Change: Turn left/right certain degrees, or direct to a navigation aid.
- Speed Change: Increase/Decrease the speed
- Readback:
- Others

The aggregate percentages of each type of communication are plotted in Fig. 5. A column in the figure represents a group of aircraft which communicated the same times. As can be seen from the figure, the transmissions about Heading Change and Altitude Change consist more than 80% of the total for the aircraft with less than 8 messages. The percentages of the inquiry messages are significantly increased when the communications times exceed 8. The increasing of inquiry communications may be the result of complex traffic situation and could also be a sign of high workload.

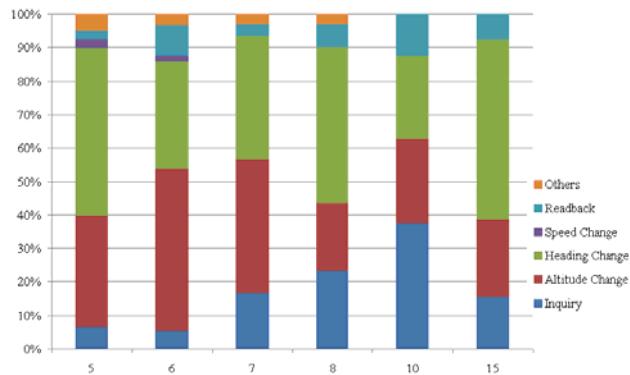


Figure 5. Distribution of different types of communication messages. The flights in all exercises are grouped into six classes according to their communication times. Each column in the figure represents a groups of flights with the number of messages illustrated below.

B. Temporal characteristics of communication activities

To determine the temporal characteristics of communication activities, we investigated the distribution of inter-communication times in each exercise.

The inter-communication time in exercise j is measured as

$$t_i^j = T_{i+1}^j - (T_i^j + L_i^j) \quad (1)$$

where T_i^j is the i^{th} communication event starts time, and L_i^j is the length of the i^{th} message.

Typically, if the data has a power law distribution $p(\tau) = \tau^{-\alpha}$, then the behavior of complementary cumulative distributions functions (CCDF) in the log-log plot will be a straight line with the slope of $-\alpha$ [27].

In practice, few empirical data obey power laws. For the most cases, data with value greater than some minimum threshold can be described in the form of power law. Here we use the method described in [28] to test and estimate the parameters of power-law, α and t_{min} . The result is shown in Table III.

TABLE III. ESTIMATED PARAMETERS FOR POWER-LAW DISTRIBUTION

	zf107	zf207	sm107	sm207	zf108	sm208
α	2.8027	2.2104	2.0097	2.4286	3.6602	2.6093
t_{min} (in seconds)	12.491	7.0141	7.2485	8.1673	19.215	15.703
proportion	0.4029	0.5213	0.6474	0.4363	0.1907	0.2513

While t_{min} specifies a lower bound of the observed data over which the data shows power-law behavior, and the proportion of each exercise describe the amount of intervals which are greater than t_{min} . However, the lack of available data to complement communication records has limited the explanation of the meaning of t_{min} .

The CCDF of inter-communication times in each exercise are plotted on log-log scale (Fig. 6). Different marker stands for the different simulation data, while the solid line is the corresponding power-law fit form. Although the behaviors of intervals in each exercise are similar, especially zf2_07, sm1_07 and sm2_07, we cannot prove that the distribution of inter-communication times follows a power-law distribution from the obtained results. More investigations on operational data should be established.

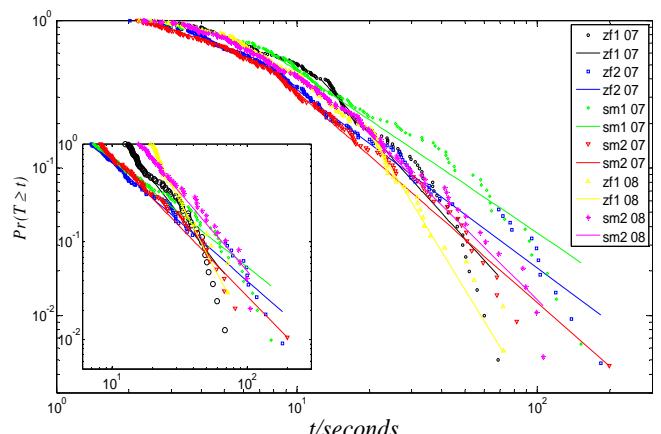


Figure 6. Distribution of inter-communication times of each exercise. The inset figure is the intervals greater than t_{min} .

V. DISCUSSION

The use of the underlying mechanisms that govern system evolution is a basic way to modeling, predicting and controlling system. While there is a great deal of expectation that the dynamics of controllers can be characterized by solely activities, our initial results demonstrate that it could not be identified by the only use of current communication data.

One of the possible reasons could be fact that the communication actions are not performed independently. For example, the sending of a next instruction may depend on the pilot's response to a previous one. Also, combining several communication instructions might be part of the air traffic controller's strategy.

Other reasons could be that the controllers were unfamiliar with the operation procedures when they did simulations. In addition, we didn't investigate the distribution of different types of communications in terms of their contents. It may be important to make this distinction as certain type of instructions, such as transfer-in and transfer-out, must be sent at a certain time.

To fully understand the complex system, we expect that not only the network theory is required to capture the emergence and structural evolution of the skeleton of the system, but also the incorporated dynamical processes that are taking place on the mentioned network[29].

This paper presented the results of our first attempt on describing controller dynamics using a human dynamics approach. We expect that in addition considering network dynamics, i.e. interconnectivities between human activities over time, would better illustrate the mechanism underlying controller dynamics. Therefore, network dynamics constitutes the core of the next steps of our investigations.

ACKNOWLEDGMENT

The authors would like to thank Prof. Patrick Bellot from Telecom-ParisTech for his advises on network dynamics. We also thank three anonymous referees for their comments and suggestions and Horst Hering for providing the ATCOSIM database.

REFERENCES

- [1] D. K. Schmidt, "On modeling ATC work load and sector capacity," *Journal of Aircraft*, vol. 13, no. 7, pp. 531-537, 1976.
- [2] D. K. Schmidt, "A queuing analysis of the air traffic controller's work load," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 8, pp. 492-498, 1978.
- [3] A. Robertson, M. Grossberg, and J. Richards, "Validation of air traffic controller workload models," *Transportation Systems Center*, Cambridge MA, 1979.
- [4] C. Manning, S. H. Mills, C. M. Fox et al., "Using air traffic control taskload measures and communication events to predict subjective workload," DOT/FAA/AM-02/4, Federal Aviation Administration, Oklahoma City, 2002.
- [5] S. Loft, P. Sanderson, A. Neal et al., "Modeling and predicting mental workload in en route air traffic control: Critical review and broader implications," *Human Factors*, vol. 49, no. 3, pp. 376-399, 2007.

- [6] G. Bedny, and D. Meister, "The Russian Theory of Activity: Current Applications to Design and Learning," Mahwah, New Jersey: Lawrence Erlbaum Associates, pp. 1, 1997.
- [7] K. Cardosi, "Time required for transmission of time-critical air traffic control messages in an en route environment," *International Journal of Aviation Psychology*, vol. 7, pp. 171 - 182, 1993.
- [8] K. Corker, B. F. Gore, K. Fleming et al., "Free flight and the context of control: Experiments and modeling to determine the impact of distributed air-ground air traffic management on safety and procedure," in *The 3rd USA/Europe Air Traffic Management R&D Seminar*, Napoli, Italy, 1997.
- [9] D. Morrow, and M. Rodvold, "Communication issues in air traffic control," *Human factors in air traffic control*, M. Smolensky and E. S. Stein, eds., pp. 421 - 456, San Diego, CA, 1998.
- [10] C. Manning, C. Fox, and E. Pfleiderer, "Relationships between measures of air traffic controller voice communications, taskload, and traffic complexity," in *The 5th USA/Europe ATM R&D Seminar*, Budapest, Hungary, 2003.
- [11] C. Manning, and E. Pfleiderer, "Relationship of Sector Activity and Sector Complexity to Air Traffic Controller Taskload," DOT/FAA/AM-06/29, Federal Aviation Administration, Oklahoma City, 2006.
- [12] J. B. Watson, "Psychology as the Behaviorist Views It," *Psychological Review*, vol. 20, pp. 158-177, 1913.
- [13] J. G. Oliveira, and A.-L. Barabasi, "Human dynamics: Darwin and Einstein correspondence patterns," *Nature*, vol. 437, no. 7063, pp. 1251-1251, 2005.
- [14] R. D. Malmgren, J. M. Hofman, L. A. N. Amaral et al., "Characterizing individual communication patterns," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, Paris, France, 2009, pp. 607-616.
- [15] U. Harder, and M. Paczuski, "Correlated dynamics in human printing behavior," *Physica A: Statistical Mechanics and its Applications*, vol. 361, no. 1, pp. 329-336, 2006.
- [16] T. Zhou, H. Kiet, J. B. Kim et al., "Role of activity in human dynamics," *EPL*, vol. 82, no. 28002, 2008.
- [17] A.-L. Barabasi, "The origin of bursts and heavy tails in human dynamics," *Nature*, vol. 435, no. 7039, pp. 207-211, 2005.
- [18] R. D. Malmgren, B. D. Stouffer, A. S. L. O. Campanharo et al., "On Universality in Human Correspondence Activity," *Science*, vol. 325, no. 5948, pp. 1696 - 1700, 2009.
- [19] R. D. Malmgren, D. B. Stouffer, A. E. Motter et al., "A Poissonian explanation for heavy tails in e-mail communication," *Proceedings of the National Academy of Sciences*, vol. 105, no. 47, pp. 18153-18158, November 25, 2008, 2008.
- [20] A. Vaquez, J. G. Oliveira, Z. Dezs et al., "Modeling bursts and heavy tails in human dynamics," *Physical Review E*, vol. 73, no. 3, pp. 036127, 2006.
- [21] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779-782, 2008.
- [22] C. Song, Z. Qu, N. Blumm et al., "Limits of Predictability in Human Mobility," *Science*, vol. 327, no. 5968, pp. 1018-1021, February 19, 2010, 2010.
- [23] J. M. Histon, R. J. Hansman, B. Gottlieb et al., "Structural Considerations and Cognitive Complexity in Air Traffic Control."
- [24] J. M. Histon, and R. J. Hansman Jr, "Mitigating complexity in air traffic control: the role of structure-based abstractions," ICAT-2008-05, ICAT, Cambridge, 2008.
- [25] K. Hofbauer, S. Petrik, and H. Hering, "The ATCOSIM corpus of non-prompted clean air traffic control speech," in *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC)*, Marrakech, Morocco, 2008.
- [26] H. Hering, "Technical analysis of ATC controller to pilot voice communication with regard to automatic speech recognition system," EEC Note N. 01/2001, EUROCONTROL Experimental Centre, Paris, France, 2001.

- [27] M. Mitzenmacher, "A brief history of generative models for power law and lognormal distributions," *Internet Mathematics*, vol. 1, no. 2, pp. 226 - 251, 2004.
- [28] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," *SIAM Review*, vol. 51, no. 4, pp. 661 - 703, 2009.
- [29] A. L. Barabasi, "The Architecture of Complexity," *Control Systems Magazine*, IEEE, vol. 27, no. 4, pp. 33-42, 2007.

Predicting Controller Communication Time for Capacity Estimation

Adan Vela, Erwan Salaün, Pierrick Burgain, William Singhose, John-Paul Clarke, Eric Feron

School of Aerospace Engineering

Georgia Institute of Technology

Atlanta GA 30332-0150, USA

avela,erwan.salaun,pburgain,singhose,johnpaul,feron@gatech.edu

Abstract—We consider the complexity and controller task load problem common to air traffic management. Expanding upon previous works that correlate controller communications to workload and complexity, a stochastic model is developed to determine the distribution of the minimum time required by an air traffic controller to manage a sector. The resulting model serves as a predictive tool for rapidly determining future workload/complexity of air traffic by considering communication time or task load time as a metric.

I. INTRODUCTION

The projected growth in air traffic demand over the next twenty years is likely to generate traffic that will exceed the control capacity of air traffic controllers under current modes of operation. Both the United States Federal Aviation Administration (FAA), and EUROCONTROL, recognize the need to predict air traffic demands for en route sectors, and plan for staffing requirements for tactical controller positions. Furthermore, it is expected that deconflicting traffic, especially at the tactical level, will require advanced decision support tools to ensure robust levels of safety [1], [2]. Consequently, there has been significant investment in the development of workload metrics to evaluate when and where capacity issues may lead to safety concerns.

The approach presented here seeks to provide an objective estimate of the probability of sector overload. Therefore, it will provide a quantitative risk analysis for capacity assessment. This study takes a statistical approach to communication capabilities based on analysis of airspace geometry and air traffic flow distributions. The work builds upon prior research on controller workload, complexity measures, and their relation to communication activity between pilots and controllers.

Controller workload is defined by Stein as “the amount of effort, both physical and psychological, expended in response to system demands (task load) and also in accordance with the operators internal standard of performance [3].” The capacity to properly manage and separate air traffic directly depends on the controller workload [4]. Unfortunately, controller workload can only be measured subjectively and depends on each individual controller’s capacity and perception. Historically, determining a sector limit capacity has relied on a simple metric: the number of aircraft present in a sector. This value is established by the Monitor Alert Parameter (MAP). The MAP value is appropriate for considering nominal traffic

patterns; however, whenever system dynamics are present (e.g. weather and changing traffic patterns), MAP values no longer accurately represent sector capacity - and often times lead to congestion, or conversely, under-utilization of the airspace.

Prior work has shown that subjective controller workload can be evaluated through objective metrics, such as complexity measures [5], [6], [7] and radio communication times [8], [9]. For example, complexity measures reflect that traffic patterns with multiple crossings and altitude changes tend to result in greater workloads for managing and separating traffic, than traffic patterns with separated trajectories. A number of factors used to evaluate complexity are based on airspace and air traffic geometry. They include aircraft density, potential conflicts, number of hand-offs between adjacent airspaces, heading and speed variation between aircraft, aircraft separation distances, and presence of weather [7], [10], [11], [12], [13], [14].

Similarly, radio communication time has been considered as an objective metric to evaluate controller workload while managing traffic. A series of experiments were conducted that concluded realistic radio activities can be used to provide objective measures of workload [15]. Additionally, another study demonstrated the high correlation between communication duration and controller workload, thereby effectively validating communication time as another workload measure [8]. Additional studies have also focused on the number and amount of communications [16], [17]. More recently, research has suggested that routine air traffic control communication events provided a good estimate of controller workload [9]. While a detailed analysis of the different type of communication events provided accurate estimates, they also concluded that the total number and duration of communication events were significantly correlated with controller workload.

This article analyzes airspace flow configuration and arrival distribution probabilities to estimate sector capacity. This approach estimates the probability distribution of minimum required communication time, which will be used as a basis for estimating the total communication time of air traffic controllers. Total communication time, which has been demonstrated to reflect controller workload, provides the theoretical framework of the proposed model. This model is based on physical considerations, and therefore includes factors used in other complexity measures. One of the advantages of this method is that it yields a probability of error or defect. In

this case, the probability that the time required to manage the airspace is beyond the allowed time interval is calculated. This is a standard criteria in operations research and may be easier to use than the subjective controller workload to assess potential air traffic safety issues.

The remainder of the paper formulates and develops a communications model to determine the distribution of the minimum required controller communication time. The formulation takes advantage of aggregate flow characteristics to generate analytical results that can be calculated in real-time. The major advancement of this paper is the introduction of a predictive communication distribution based on estimated traffic flow into a sector. Furthermore, the model can be adjusted in real-time for dynamic analysis of any en route airspace. Additionally, the model is inherently robust in determining sector task load for establishing sector complexity. While, dynamic density and other metrics require strict knowledge of aircraft positions and intentions for establishing complexity, this model does not require such information. As such, it provides a framework for establishing capacity and fragility of the system based on probability distributions in a manner that is effective for long-term planning by traffic flow managers.

Section II provides a general description of the problem. Section III describes the general mathematical process describing controller communication. In Section IV, an en route sector is analyzed, resulting in a mathematical model describing the distribution properties for the minimum standard communications required to manage traffic. A case study is presented in Section V. Then, in Section VI and in Section VII, our future works and conclusions are presented.

II. PROBLEM DESCRIPTION

Consider a set of aircraft flows, $\mathcal{F}_1 \dots \mathcal{F}_n$ traversing a sector, as shown in Fig. 1. The majority of flows consist of commonly utilized jet routes that exist in \mathbb{R}^3 . Such dominant aircraft flows may consist of planer tracks, or trajectories that ascend or descend in altitude. Furthermore, tracks may consist of a series of intersections, merge points, and splits. In general, each flow is defined according to its entrance and exit locations, altitude, and designated way points. The aircraft arrivals into each flow are shown to occur according to a stochastic process in Section IV. All these flow characteristics (i.e. route and arrival process) can be determined from ETMS (Enhanced Traffic Management System) data, as shown in [18].

A communications model will be developed according to real-world operations. At a minimum, each aircraft is communicated with at least twice by the managing air traffic controller, once to acknowledge the aircraft as it enters the sector, and again when the aircraft leaves as part of the hand-off to the next sector. Additionally, pilots will acknowledge and read-back any communications from the air traffic controller. Depending on weather conditions (including turbulence), additional messages may be passed to pilots. In response to such messages, pilots will typically request clearance to fly at another flight level or to propose a new route, to which controllers will permit or suggest other options.

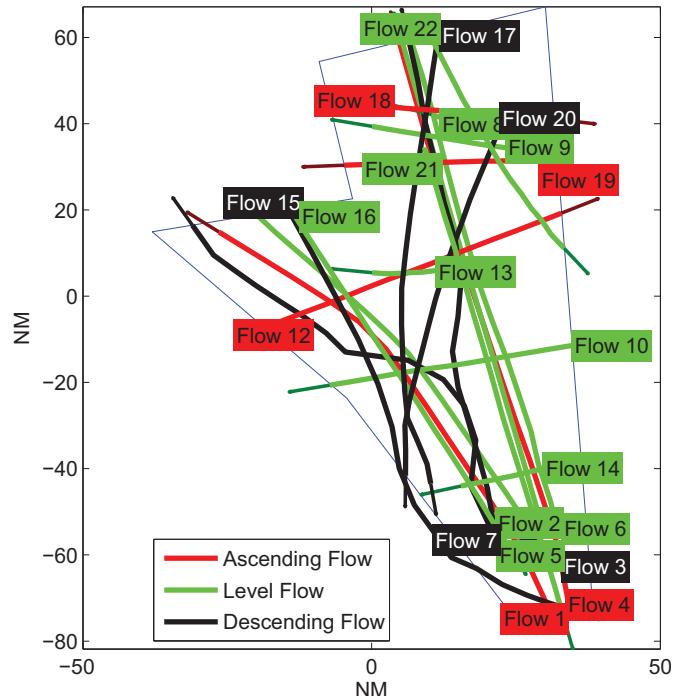


Fig. 1. Sector Map of ZTL36 [NM] with common flows indicated

Another prevalent communication typically occurs when an airspace is congested, and there is the potential for conflict or proximity. Aircraft are considered to be in proximity when there is sufficient concern for air traffic controllers to check for conflict and issue resolution commands. When conflicts are present, air traffic controllers must determine safe routes for all aircraft, and communicate these to each pilot. For this process to occur in a safe manner there must be sufficient time for the controller to gain situational awareness, determine a course for the aircraft, dictate any resolution commands, and then monitor implementation of the commands. However, as will be explained further, the probability of conflict between aircraft is not explicitly taken into account in this article.

For each event (arrival, departure, conflict) there will be an associated stochastic model describing that process. Corresponding to each event is also a time cost. For example, each aircraft entering from \mathcal{F}_i will initiate communication with the controller through a spoken message. That message will then be followed by a controller response, and possibly additional communications depending on environmental factors within the airspace. As such, the total communication time required for the interaction is given by the random variable $T_{i,a}$, where a denotes the arrival process and i is the flow. Similarly, each other event will have a corresponding time. For the case of conflict communications, this is in itself a random process. Each aircraft entering the sector will initially be conflict-free. However, due to traffic configurations a fraction of aircraft entering into the sector will require conflict resolution and communication. While behavior of air traffic controllers depends on the congestion within the sector, for the

current model presented, the time required to generate a safe resolution and to communicate the action will be considered to be fixed.

The input to the problem will be a description of the total aircraft arrival and departure rate into and out of the airspace, λ_a and λ_d , as well as a description of the traffic probability distributed among the flows, $[p_1, \dots, p_n]$, and any distinctive features of the flow trajectories (crossing angles, altitude changes, etc). The output will be a probability distribution for all required standard communications, i.e. the total minimum required communication time, \mathbf{T}_m , corresponding to the summation of each communication time required.

The minimum required communication time, \mathbf{T}_m , will be considered over a 1 minute period. Communication with aircraft entering a sector can be delayed for a short time period, as long as no immediate conflicts are present. Following a 1 minute communication, aircraft typically traverse between 5-10NM, so conflicts are unlikely to be immediate. The work can be extended to 5 minute, 10 minute, and 20 minute periods. The longer time periods establish controller task load over extended periods of time. In particular, it is acknowledged that air traffic controllers are able to manage high-levels of workload over short periods of time, however, this is typically difficult to maintain for longer periods of time. In regards to the communications model, the maximum allowed total time spent on communications determined by controller workload limits, decreases over the length of the time period considered. That is, if $\bar{T}(5)$ and $\bar{T}(20)$ correspond to the maximum total task load of an air traffic controller over 5 minutes and 20 minutes, than we should expect that $4 \cdot \bar{T}(5) \geq \bar{T}(20)$.

III. MINIMUM REQUIRED COMMUNICATION TIME

The proposed model considers the minimum communication time required by an air traffic controller to minimally manage traffic (i.e. with no conflict). Communication time of the controller is segmented according to task: acknowledgement of aircraft entering the center; clearance and requests for aircraft requiring altitude or other trajectory changes; and notice to departing aircraft of the frequency in the next airspace. There are also additional communications that may occur: courtesy statements, advisories (weather, traffic, etc), and repetitions of commands.

Each communication type is associated with an interval of time, during which time no other tasks or events can occur (mental or vocal). The time intervals take into account all interchange between pilots and the controller, including pauses. Let the random variable representing the total time durations for each task be defined accordingly:

- Acknowledgement: \mathbf{T}_a
- Clearances and requests: \mathbf{T}_c
- Frequency changes: \mathbf{T}_f .
- Repetitions: \mathbf{T}_r

Other forms of communication will be left out for now. Particularly, advisory statement will not be initially included in the communications, as this is typically a discrete mode of operation. For example, if weather, turbulence, or traffic

congestion is present, then such information will be repeatedly relayed to all aircraft by the controller. As such, if there exist environmental conditions that warrant an advisory, then the additional time can be included as part of the total acknowledgement time, \mathbf{T}_a . Furthermore, define \mathbf{T}_c to be sum of the minimum time required for a request and clearance communication to take place. This occurs when the airspace is relatively clear such that the probability of conflict is low. The time required for hand-off of aircraft between airspaces is \mathbf{T}_f . And \mathbf{T}_r is the additional time required to repeat any communications.

From the definitions above, the random variable for minimum required communication time, \mathbf{T}_m within any 1 minute time block is given by:

$$\mathbf{T}_m = \mathbf{T}_a + \mathbf{T}_c + \mathbf{T}_r + \mathbf{T}_f. \quad (1)$$

Additionally, we will make use of random variables for the number of events for each communication, to be labeled by \mathbf{N}_a , \mathbf{N}_c , \mathbf{N}_r , and \mathbf{N}_f corresponding to the previously defined time duration blocks.

The value of \mathbf{T}_m will refer to the minimum standard communication expected between pilots and the air traffic controller. This is a corollary of the assumption that request and clearance communications only occur for the simple conflict-free cases. The present model lacks a notion of conflict detection and resolution. As such, given that \mathbf{T}_m is the minimum time required for resolving traffic, then it is asserted that $\mathbf{T}^{free} = 1 - \mathbf{T}_m$ is free time available for resolving conflicts, and issuing clearances for more difficult configurations.

IV. AIRSPACE MODELING: ARRIVALS AND DEPARTURES

To accurately generate a model of standard communications, it is necessary that the traffic flow into the sector is appropriately described from historical data [18]. From such a description, the arrival rate of communication events can be mathematically characterized. In this section, statistical information on aircraft arrivals into, and departures out of, a sector will be provided. It will be demonstrated that the event counting will follow a Poisson counting process, with the associated exponential inter-arrival times. A Poisson model is beneficial as it allows for analytical modeling of the problem, and appears to accurately reflect aircraft events for long-term planning.

For an arbitrary counting process, \mathbf{N} , defined by parameter λ [events/minute], the probability distribution is given by [19]:

$$P(\mathbf{N}(t) - \mathbf{N}(t-s) = k) = \frac{(\lambda)^k}{k!} e^{-\lambda s}, \quad (2)$$

with the associated cumulative distribution for the random variable \mathbf{I} representing event inter-arrival times, is given by:

$$P(\mathbf{I} < a) = 1 - e^{-\lambda a}. \quad (3)$$

To drive the modeling process and provide verification of the event model in (2), the airspace ZTL36 (near Atlanta, GA, USA) was selected for analysis as a sample case. The sector is generally a high-complexity sector with many flows, as it

is dominated by north-bound ascending aircraft from KATL airport, and both east-bound and west-bound crossing traffic at the northern and southern ends of the sector. There are also other less prevalently traversed paths that occur spatially near dominating flows that lead to increased complexity. A figure of the sector geometry with clustered flights is shown in Fig. 1. A 15 minute time period from 20:00-20:15 GMT (i.e. 2pm-2:15pm EST) over 42 days was selected for analysis, where the dates are sampled from June 12-August 22 of 2005. The selection of the time period is arbitrary, however, for the case at hand, a high traffic period was present during these hours. Average, maximum and minimum bounds on aircraft counts within the sector over time are shown in Fig. 2. Over a 24hr time period, it is clear that traffic demand into the sector fluctuates; the varying demand is noticeable due to the proximity of the sector to KATL. Therefore, it is expected that the arrival and departure process of aircraft is non-homogeneous. However, for 15 minute time periods, both the arrival and departure process can be modeled as homogeneous. It follows then that the event counting process model for any 1 minute period, can be modeled according to a 15 minute interval centered about the 1 minute period in question.

For the 15 minute period considered, the sampled cumulative distribution function of arrival times is shown in Fig. 3. Also, shown in Fig. 3 is the departure sampled cumulative distribution function (CDF) during the same time period. The inter-arrival times of aircraft arriving into and departing from the sector fits an exponential distribution (3), with parameters $\lambda_a \sim 1.25$ [aircraft/min] and $\lambda_d \sim 1.10$ [aircraft/min], respectively. Modeling air traffic arriving into a sector as a Poisson process allows for simple modeling of the system.

Note that the arrival and departure distributions are not the same. Given that the distribution of aircraft service times through the sector, that is the time to traverse the sector, ranges from 5-30 minutes, the departure rate can lag the arrival rate. If the arrival process is homogeneous, $\dot{\lambda}_a(t) = 0$, then the arrivals and departures processes will have the same distribution. It appears that at least for sectors near major airports, it is unlikely arrival process will be homogeneous over 15-45 minute periods due to dynamic demand in the local area as a result of the airport.

For example, inter-arrival times for the set of aircraft entering the sector that request altitude clearances can be modeled as exponential. This assumption is validated for all the major flows present. Figure 4, for example, illustrates that aircraft at level flight entering the sector along flow 5 are distributed according to an exponential distribution like in (3). Related work on sector modeling for ZTL19 is provided in [18]. Specifically, through the Kolmogorov-Smirnov hypothesis testing it is demonstrated that aircraft arrivals and departures are Poisson for the majority of the day, the only exception being during low demand time periods (e.g. 4AM-10AM [GMT], 10PM-5AM local time) [18]. From the above description of the traffic arrival and departure model, we will make the assumption that flows into the sector are also Poisson, and inter-arrival times are described through an

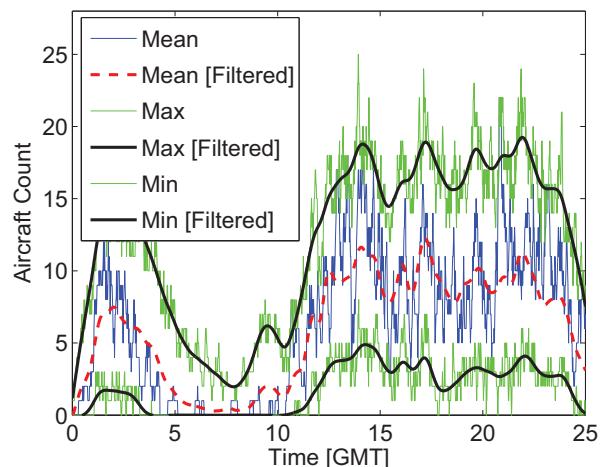


Fig. 2. Aircraft counts in ZTL36 over 24hr period

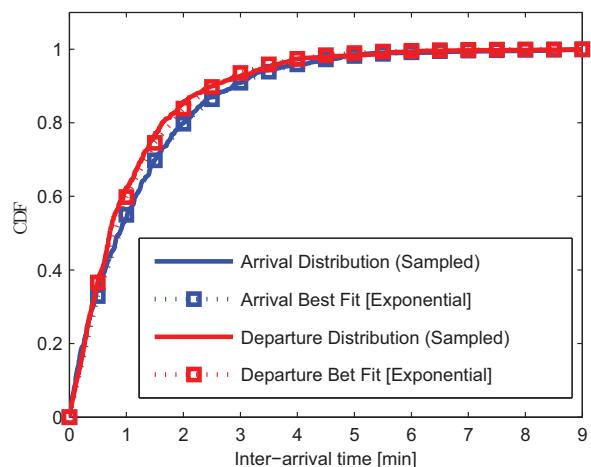


Fig. 3. CDF of aircraft inter-arrival time into the sector between 20:00-20:15 GMT

exponential distribution.

It is now possible to describe the distribution of communication time in a form similar to (1). The arrival process will be broken into multiple processes. Let the random variable N_a represent the total number of aircraft that arrive within the 1 minute period. The total can be subdivided according to flows and required controller actions. The counting process for each flow \mathcal{F}_i is grouped and divided into four sets:

$$\begin{aligned} S_1 &= \{i | \mathcal{F}_i \text{ is at level flight and requires repetition}\} \\ S_2 &= \{i | \mathcal{F}_i \text{ is at level flight and does not require repetition}\} \\ S_3 &= \{i | \mathcal{F}_i \text{ requires clearance and repetition}\} \\ S_4 &= \{i | \mathcal{F}_i \text{ requires clearance and no repetition}\}. \end{aligned} \quad (4)$$

For example, S_1 is the set of all aircraft arrivals that are at level flight, and require repetition of any acknowledgements between the controller and pilot.

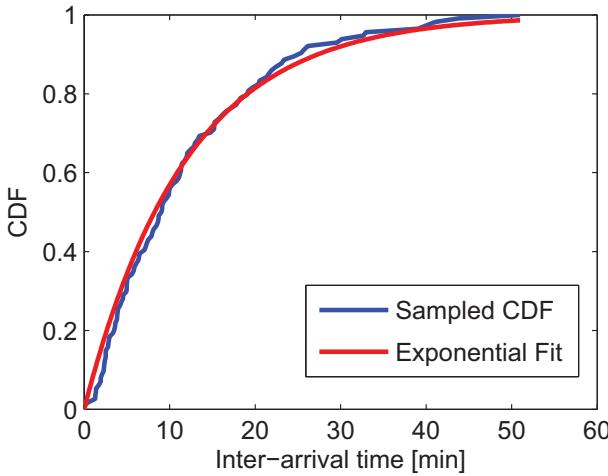


Fig. 4. CDF of aircraft inter-arrival time into flow 5 between 20:00-20:15 GMT

Then,

$$\mathbf{N}_a = \mathbf{N}_1 + \mathbf{N}_2 + \mathbf{N}_3 + \mathbf{N}_4, \quad (5)$$

where, each random variable \mathbf{N}_k is the associated counting process for the set S_k listed in (4).

The λ_k parameter characterizing the counting process \mathbf{N}_k in (5) can be calculated using flow parameters, and the probability distribution of repetition being required. If $\lambda_{i,a}$ is the mean arrival rate for flow \mathcal{F}_i , and the probability of aircraft arriving into the sector requiring repetition is $p_{i,a,r}$, then the rate of aircraft from flow \mathcal{F}_i requiring repetition is $\lambda_{i,r} = p_{i,a,r}\lambda_{i,a}$. Similarly, this can be calculated for all other flows. For each set in (4) the arrival parameter is given by:

$$\begin{aligned} \lambda_1 &= \sum_{i \in S_1} p_{i,a,r}\lambda_{i,a} & \lambda_3 &= \sum_{i \in S_3} p_{i,a,r}\lambda_{i,a} \\ \lambda_2 &= \sum_{i \in S_2} (1 - p_{i,a,r})\lambda_{i,a} & \lambda_4 &= \sum_{i \in S_4} (1 - p_{i,a,r})\lambda_{i,a}. \end{aligned}$$

The compound Poisson process for the above arrival models is in fact the original aircraft arrival model when only considering aircraft arrivals into the sector.

The counting process for departure events can be expressed similarly, defining the sets

$$\begin{aligned} S_5 &= \{i | \mathcal{F}_i \text{ departing requiring repetition}\} \\ S_6 &= \{i | \mathcal{F}_i \text{ departing not requiring repetition}\} \end{aligned} \quad (6)$$

with corresponding counting process, \mathbf{N}_5 and \mathbf{N}_6 defined by parameters:

$$\lambda_5 = \sum_{i \in S_5} p_{i,d,r}\lambda_{i,d} \quad \lambda_6 = \sum_{i \in S_6} (1 - p_{i,d,r})\lambda_{i,d},$$

where $\lambda_{i,d}$ is the departure rate for flow \mathcal{F}_i , and $p_{i,d,r}$ is the probability departing aircraft from flow \mathcal{F}_i require repetition.

Each Poisson process S_k is associated with a constant communication time, T_k . For example, all aircraft arriving into the sector requiring clearance and repetition will require T_3

seconds of communication between the pilot and the controller. The random variable for total communication time \mathbf{C}_k required for each set S_k is

$$\mathbf{C}_k = \mathbf{N}_k T_k \quad k = 1, \dots, 6,$$

and the probability distribution for each \mathbf{C}_k is:

$$f_k(c) = P(C_k = c) = \begin{cases} \frac{\lambda_k^{c/T_k}}{(c/T_k)!} e^{-\lambda_k}, & c/T_k \in \mathbb{Z}^{0,+} \\ 0 & \text{else.} \end{cases}$$

The random variable for the total minimum standard communication time required of a controller in the minute is:

$$\mathbf{T}_m = \sum_{k=1}^6 \mathbf{C}_k.$$

The final result for the distribution of \mathbf{T}_m can be determined analytically through the convolution:

$$P(\mathbf{T}_m = c) = f_1(c) \otimes f_2(c) \otimes \dots \otimes f_6(c).$$

V. CASE STUDY

The case study presented below considers the minimum required standard communications model.

Consider the sector ZTL36 as previously described between the time window 20:00-20:15 [GMT]. According to historical data, the aircraft arrival rate into the sector is $\lambda_a = 1.25$ [aircraft/min], while the aircraft departure rate is $\lambda_d = 1.10$ [aircraft/min]. The following aircraft arrival rates and aircraft departure rates are calculated according to the sets defined in (4) and (6):

$$\begin{aligned} \lambda_1 &= p_{a,l} p_{a,l,r} \lambda_a & \lambda_5 &= p_{d,r} \lambda_d \\ \lambda_2 &= p_{a,l} (1 - p_{a,l,r}) \lambda_a & \lambda_6 &= (1 - p_{d,r}) \lambda_d \\ \lambda_3 &= (1 - p_{a,l}) p_{a,c,r} \lambda_a & \\ \lambda_4 &= (1 - p_{a,c}) (1 - p_{a,c,r}) \lambda_a & \end{aligned} \quad (7)$$

where $p_{a,l}$ is the probability an aircraft is at level-flight; $p_{a,l,r}$ and $p_{a,c,r}$ are the fraction of arrivals at level-flight and arrival requiring clearance that will have read-back; and $p_{d,r}$ is the probability an aircraft departing the sector will require read-back. According to historical data for the time period considered, the percentage of arriving aircraft at level flight is $p_{a,l} = 0.54$. The percentage of communications requiring repetition is taken to be $p_{a,l,r} = p_{a,c,r} = p_{d,r} = 0.05$. The percentage of communications requiring repetition is based on sampled air traffic controller communications. (Out of 109 communications between pilots and controllers, there were 6 communications that required repetition or clarification.)

The time required for each communication is listed in Table I. For example, the communication time for arriving aircraft requiring clearance under standard communication is $T_4 = 11$ s. However, if repetition is required, then the total communication time required required is $T_3 = 15$ s. The communication times above are averages based on a sample audio recording of air traffic controllers within the ZTL for

TABLE I
AVERAGE COMMUNICATION TIME FOR EACH EVENT (SECONDS).

$T_1 = 10$	$T_2 = 6$	Arrivals at level-flight
$T_3 = 15$	$T_4 = 11$	Arrivals requiring clearance
$T_5 = 8$	$T_6 = 6$	Departure

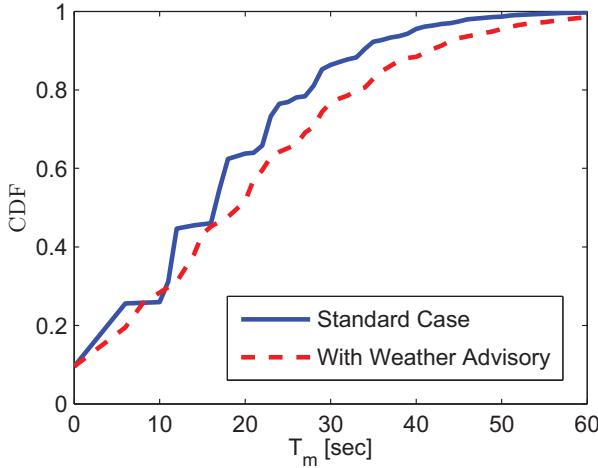


Fig. 5. Example distribution for minimum expected communication time over 1 minute

November 3, 2010 over a 1 hour time period.

Based on the aircraft arrival rate into the sector, and the departure rate of aircraft out of the sector, as well as the communication times, the cumulative distribution of the minimum required communication time is shown in Fig. 5. In the case of a weather advisory to all aircraft, the effect is to increase the initial communication times for all entering aircraft. For each communication event relating to arrivals, additional time is required for air traffic controllers to relay the additional information to pilots. In this case, $T'_k = T_k + 2$ for $k = 1, 2, 3, 4$, where 2 seconds is the additional time required. For example, typically aircraft arriving at level flight required 6 seconds of communication. If weather is present, then aircraft now require 8 seconds of communication. The effect of a weather advisory on the cumulative distribution of the minimum required communication time is also shown in Fig. 5. As expected, the probability of longer communication times increases.

This work can be expanded to include the probability that air traffic controllers will see difficult workload conditions in any period. Assuming that 45 seconds is an upper threshold for minimum standard communications by a controller, Fig. 6 shows the probability a pair of aircraft arrival and departure rates will lead to high levels of workload.

VI. FUTURE WORK

The work presented thus far concentrates on the minimum required communication time for air traffic controllers. However, this is clearly an incomplete model, as it does

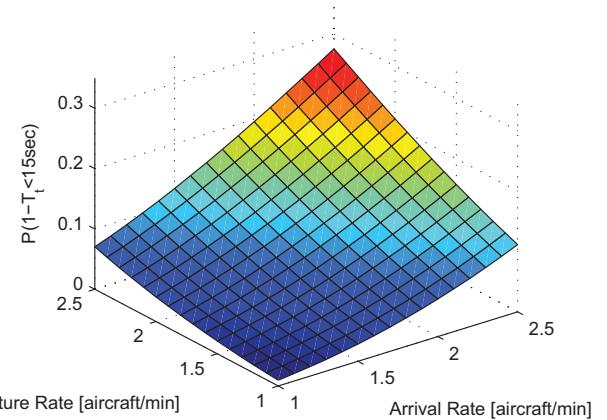


Fig. 6. Probability of standard communication time requiring more than 45 second.

not consider task load related to potential aircraft conflicts or proximity of aircraft that requires additional controller intervention and monitoring.

To be a complete task load model, we must consider additional tasks related to aircraft in proximity or conflict. Aircraft conflicts require significant time commitment from controllers in the form of:

- Assessing and evaluating possible conflicts
- Determining conflict resolution solutions
- Communicating resolutions to pilots
- Monitoring implementations of resolution commands

It is important to note, that the majority of the tasks and effort involved with potential conflicts do not revolve around the communication of the solution from controller to pilot. In fact, much of the task involves assessing and solving potential conflicts. Additionally, monitoring aircraft to ensure resolution commands are implemented requires significant controller effort.

Define N_c to be the number of conflict events that occur within any given time period. Furthermore, let $T_x = T_x N_x$ be the total time involving the conflict resolution process, according to the tasks listed previously. Then the total controller task load time, T_T , can than be established as follows:

$$T_T = T_m + T_x. \quad (8)$$

The total task load time T_T could provide a more complete model of predictive controller task-load. Future work, will develop a conflict model to establish an event arrival process for conflicts and proximity events.

Finally, addition work is required on establishing communication times for each type of event through greater sampling of air traffic communications.

VII. CONCLUSION

A new method for predicting controller task load according to communication time has been presented. The result is

a stochastic process model based on aircraft arrivals and departures, which considers the underlying counting processes. From the stochastic model, it is possible to predict in real-time a metric of the workload an air traffic controller will expect. Furthermore, because the model considers random distributions, it is possible to establish confidence bounds on the predicted task load. As such, the model can be used as a predictive tool for traffic flow managers to adjust traffic flows to be within controller limits. A major result of the paper is the introduction of the sensitivity of the controller task load. By establishing a task load distribution, it is possible to determine the probability the system will overwhelm a controller.

VIII. ACKNOWLEDGMENTS

This work is funded by NASA under Grant NNX08AY52A and by the FAA under Award No.: 07-C-NE-GIT, Amendment Nos. 005, 010, and 020. Additionally, the authors would like to acknowledge and thank Vlad Popescu, Jessica Derenzy, and Rachel Haga for providing annotated data relating to controller communications.

REFERENCES

- [1] NextGen Integration and Implementation Office, "Nextgen implementation plan," Federal Aviation Administration, Tech. Rep., 2009.
- [2] J. Villiers, "En route air traffic soft management ultimate system," *Institut du Transport Aerien*, vol. 58, 2004.
- [3] E. Stein, *Human Operator Workload in Air Traffic Control*, S. E. Smolensky MW, Ed. New-York: Academic Press, 1998.
- [4] P. Brooker, "Control workload, airspace capacity and future systems," *Human Factors and Aerospace Safety*, vol. 3, no. 1, pp. 1–23, 2003.
- [5] P. Kopardekar, J. Rhodes, A. Schwartz, S. Magyarits, and B. Willems, "Relationship of maximum manageable air traffic control complexity and sector capacity," in *Proc. of the 26th International Congress of the Aeronautical Sciences*, Anchorage, AK, 2008.
- [6] S. Athenes, P. Avery, S. Puechmorel, D. Delahaye, and C. Collet, "ATC complexity and controller workload: Trying to bridge the gap," in *Proc. of the International Conference on Human-Computer Interaction in Aeronautics*, Cambridge, Massachusetts, 2002.
- [7] B. Arad, "The control load and sector design," *Journal of Air Traffic Control*, pp. 13–31, 1964.
- [8] D. Porterfield, "Evaluating controller communication time as a measure of workload," *International Journal of Aviation Psychology*, vol. 7, no. 2, pp. 171–182, 1997.
- [9] C. Manning, S. Mills, C. Fox, E. Pfleiderer, and H. Mogilka, "Using air traffic control taskload measure and communication events to predict subjective workload," Federal Aviation Administration, Office of Aerospace Medicine, Tech. Rep., 2002.
- [10] G. Chatterji and B. Shridhar, "Measures for air traffic controller workload prediction," in *Proc. of the 1st AIAA Aircraft Technology, Integration, and Operations Forum*, Los Angeles, California, 2001.
- [11] P. Kopardekar and S. Magyarits, "Measurement and prediction of airspace complexity," in *Proc. of the 5th US/Europe Air Traffic Management Seminar*, Budapest, Hungary, 2003.
- [12] D. Delahaye, S. Puechmorel, J. Hansman, and J. Histon, "Air traffic complexity based on non linear dynamical systems," in *Proc. of the 5th USA-Europe ATM R&D Seminar*, Budapest, Hungary, 2003.
- [13] I. Laudeman, S. Shelden, R. Branstrom, and C. Brasil, "Dynamic density: An air traffic management metric," *Ames Research Center-112226*, 1998.
- [14] M. Prandini, V. Putta, and J. Hu, "A probabilistic measure of air traffic complexity in three-dimensional airspace," *International Journal of Adaptive Control and Signal Processing*, vol. 1, pp. 1–25, 2009.
- [15] C. A. Shingledecker, "Subsidiary radio communications tasks for workload assessment in R&D simulations," Air force aerospace medical research laboratory, Tech. Rep., 1982.
- [16] K. Corker, B. Gore, K. Fleming, and J. Lane, "Free flight and the context of control: Experiments and modeling to determine the impact of distributed air-ground air traffic management on safety and procedures," in *Proc. of the 3rd Annual Eurocontrol International Symposium on Air Traffic Management*, Naples, Italy, 2000.
- [17] K. Cardosi, "An analysis of en route controller-pilot voice communications," US Department of Transportation, Federal Aviation Administration, Tech. Rep. DOT/FAA/RD-93/11, 1993.
- [18] E. Salaün, M. Gariel, A. Vela, E. Feron, and J.-P. Clarke, "Statistical proximity maps based on data-driven flow modeling," in *Proc. of the AIAA Infotech@Aerospace*, Atlanta, GA, 2010.
- [19] S. Ross, *Stochastic processes*, 2nd ed. Wiley series in probability and mathematical statistics, 1996.

Track 9

Prospective Studies and Economics

What Kind of Aviation Infrastructure Privatization is Needed in China?

Wei Lu

College of Economy and Management
 Civil Aviation University of China
 Tianjin 300300, P.R.China
 Visiting Associate Professor at EPFL
 wei.lu@epfl.ch

Matthias Finger

Management of Network Industries
 EPFL - Ecole Polytechnique Fédérale Lausanne
 Lausanne 1015, Switzerland
 Matthias.finger@epfl.ch

Abstract—Aviation infrastructures used to be considered critical to national security and the public interest. They were also considered natural monopolies. Consequently, it was believed that government or public entities should be responsible for the ownership and management of aviation infrastructures. However, since the late 1970s, commercialization and privatization began to become increasingly widespread in airports. This paper will investigate underlying rationales for the introduction of private sector participation in aviation infrastructures, be it in terms of privatization or in terms of delegated management, and all this both in the cases of China's and developed economies' airports and air traffic management. It is argued that partial privatization may be much more appropriate in the case of China's aviation infrastructure sector.

Keywords- *aviation infrastructure; privatization; china*

I. INTRODUCTION

Air transportation is a fast growing sector and is forecasted to keep on growing in the future. Airbus (2009) suggests that overall world passenger traffic is expected to increase by 4.7% per annum and the numbers of frequencies offered on passenger routes will more than double during 2009-2028. This also means that more flights and aircrafts are needed so as to accommodate those increases. But one critical problem for the aviation industry lies in the fact whether sufficient and qualified infrastructure services could be provided so as to accommodate those increases. Nowadays, most of the aviation infrastructures, i.e., airport and air traffic management (ATM), are provided by government or some other public entities. And there exist controversies in many countries about how to operate and develop these infrastructure services (Button and McDougall 2006).

II. REGULATIONS IN AVIATION INFRASTRUCTURE SECTORS: RATIONALES AND PRACTICES

Although most industries are regulated to varying degrees, few of them are regulated as heavily as the aviation industry (Vasigh, Tacker et al. 2008). This is especially the case for the infrastructure aspect of the industry. Aviation infrastructures used to be managed by governments or public organizations around world. There are several rationales in favor of public management of aviation infrastructures.

A. Theory of natural monopoly

According to natural monopoly theory, industries with such characteristics would enjoy the benefit of economies of scale, scope and density (Chen, Tan et al. 2004). Hence, it is economically efficient that only one firm exists in the market. It is believed that both the airport and ATM segments are natural monopolies, in which only one service provided is required for the optimal economic efficiency. More providers can only increase average costs for services provided. Hence, monopolistic airport and ATM service provision is optimal so as to achieve lowest service cost.

B. Public interest consideration

As for monopolistic markets, one is generally worried about the possibility of monopolistic firm's abuse of market power. It is believed that, compared with private firms which set profit maximization as their ultimate objective, public organization would take more responsibility as to public interests. Especially referring to the aviation industry, it is argued that simply emphasizing business practices could cause damage to certain public interests. For example, if airports allocated their slots simply based on profit maximization considerations, regional carriers and general aviation users would find it difficult to compete with network legacy carriers and would lose many slots allocated to them. Hence, it is argued that public organizations would better consider public interests.

C. Worries about service quality and safety standards after privatization

One incentive for privatization lies in the fact that cost optimization associated with privatization could reduce production cost and make organizational operation more efficient. But referring to the aviation industry, as safety is the most important factor for the sustainability of the industry, there exist worries that cost reduction introduced by privatization would cause a lowering of service quality and safety standards (Donohue 1999).

Based on such considerations, government tends to use public service providers to operate airports and/or ATM infrastructures and supply services herewith. As to airports, it

used to be government or some other public entity that provided airport services. A survey conducted by ICAO in 1999 shows that “the large majority of airports remain under government or public ownership either wholly or through a majority holding” (World Trade Organization 2006, p.91), although airport privatization became much more popular in the sector.

ATM service has a much higher tradition of being managed by government or public entities. Up to now, there is only limited privatization (or as commonly labeled as “commercialization) in this regard. Except for the rationale for natural monopolies, proponents of privatization also argue that air traffic management is critically important for overall system safety and that a private party would sacrifice ATM safety standards for profitability.

III. PRIVATIZATION DEVELOPMENTS IN AVIATION INFRASTRUCTURES

Since the late 1970s when deregulation began in the aviation industry, there are more and more cases of airports and/or air traffic management organizations privatization or at least conversion into businesslike organization. But the term privatization remains ambiguous. Vickers (2008) suggests that, “privatization is the transfer from government to private parties of the ownership of firms”. Even though in aviation infrastructures this definition seems quite unrealistic, U.K. style airport privatization follows this very idea. The majority of airport privatization only takes forms of partial privatization or even of no ownership transfer. For example, in the case of the U.S. airports, there are few examples in which government or other public entities try to sell their airport ownership to private parties. As for the ATM sector, there exists no full privatization at all. Therefore, it is much more realistic to define aviation infrastructure privatization from a board sense and abandon the effort to simply base aviation infrastructure privatization on ownership structure. In this paper, it is suggested that aviation infrastructure privatization should be considered as the transfer of some degree of control from government to private or businesslike entities.

With the development of economic theories and industry practices, it is argued that the past understanding of a natural monopoly may not be appropriate. Some economists argue that past entry regulation for natural monopoly industry is not necessary, as potential competitors could help to increase economic efficiency of the incumbent supplier (Chen, Tan et al. 2004). Especially there exists significant space for competition in the airport segment. It can be seen that some airport services like catering and ground-handling services, lift of entry limitation always lower charges concerned and increase the service quality provided. This was also the rationale for the EU's initiatives to reform its ground-handling market since 1993.

Another incentive for aviation infrastructure privatization lies in the soaring demand for air service, which also increases the demands for aviation infrastructure services significantly. Confronted with fast increasing air service demand, governments or public organizations responsible for airports and air traffic management find that they lack the flexibility to

generate sufficient revenue needed for aviation infrastructure improvement. As there exist complicated regulation and procedures regarding to public expenditure, investment in aviation infrastructure always lags behind industry demand. For example, there are several independent commissions report that the infrastructure for air traffic management in US is deteriorating and has an inadequate source of capital funds needed to modernize (Donohue 1999). Also, congestion in many hub airports simply implies that more efficient ATM services, together with airport services, are needed for the sustainable operation of the airline industry. All those theoretical and practical developments are incentives for privatizing aviation infrastructures.

Up to now, most of the aviation infrastructure privatization cases happen in the airport segment. The benchmark for airport privatization is the U.K. model. Driven by the Thatcher government's national privatization campaign, the Airports Act was passed in 1986. Under this act, the U.K. government transformed the British Airport Authority into the private BAA plc and all other airports with a turnover of more than £ 1 million were required to be corporatized. By 2007, all main airports in the U.K. (except the Manchester airport group) were transferred to private ownership and most are 100% privately held (Graham 2008). Since the late 1990s, Australia and New Zealand have also changed the ownership of their major airports to private ownership, a move similar to that of the U.K. (Forsyth 2008).

While in the airport segment, U.K. style full airport privatization is quite radical for most countries, even US and other European countries do not follow this full privatization strategy. U.S. airports are primarily publicly owned and belong to local governments. The US privatization approach has been quite different from the dominant privatization doctrine, as there were generally no ownership transfers to private entities. Neufville and Odoni (2003) argue that U.S. airports can nevertheless be considered to be the “most privatized airports”, as airport operators extensively outsource most of their airport business and operational functions to private entities. Also, many airports in Europe are still owned by national, state, or local government. However, and although there was no full airport privatization like in the U.K., the governance structure and business orientation of several public European airports has been changed and become more profit oriented (Gillen and Niemeier 2008).

Airport privatization is also prevailing in many developing countries. Those countries have followed a proactive, yet cautious approach when transferring airport ownership to private entities. From 1990-2008, 47 developing countries introduced private participation programs granting 132 airport programs. Latin America and the Caribbean countries were the pioneers of airport privatization among these countries (World Bank 2009). Concessions were the predominant type of airport privatization in Latin America. For instance, in 1998 the Argentine government transferred 33 of the country's total 59 airports to private Aeropuertos Argentina 2000 for the duration of 30 years, with the total investment amounting to 2.2 billion US\$ (Lipovich 2008).

Compared with the airport segment, the privatization efforts in ATM occurred much later and was more limited. Up to now, there is not even though any example of full privatization in this segment. Even there are no cases of ATM privatization, there exist a variety of institutional arrangement, including U.K.'s public private partnership arrangement, Canada's "not for profit" private corporation, together with other countries' government corporation or "not for profit" joint-stock corporation (Goodliffe 2002; Button and McDougall 2006). Due to the difficult of quantitative measurement of ATM performance and different standards applied in different system, it is difficult to measure efficiency improvements after the introduction of those institutional changes in the ATM sector. But Button and McDougall (2006) suggests that some commercialized ATM provider reduced their cost per instrument flight rules movement on average by about 15% during 1997-2004, while state-owned FAA had an increase of 23% in the same period¹. They also indicate that ATM providers become more responsive to user demands. Another research done by the US General Accounting Office (GAO) suggests that data from the five air navigation service providers indicate that since commercialization, the safety of air navigation services has remained the same or improved (United States. Government Accountability Office 2005)

IV. DEVELOPMENTS IN CHINA'S AVIATION INFRASTRUCTURES

Due to the fast paced economic expansion, especially after 1992 when China began its transition from a planned economy to a market economy, the air transportation market has maintained a rapid growth pace. From figure 1 it can be seen that China's civil aviation market has sustained an average yearly growth rate of 16% during 1985 – 2008. According to International Civil Aviation Organization (ICAO)'s statistics, since 2005, China's total air transportation turnover (not including Hong Kong, Macau and Taiwan) has been ranked second only to that of the US. China's fast growing air transportation market places great pressure upon airport operators and air traffic management service providers (Civil Aviation Administration of China 2008).

With fast growing air service demands, the gaps between China's aviation infrastructure capacity and soaring air service demands also increase rapidly. In the late 1970s China had about 70 civil airports, and this number had increased to 152 by 2008 (Zheng, Lu et al. 2009). In the past three decades, about 80 new airports have been built and significant improvement projects have been introduced to the existing ones. Yet, for the world's second largest air transportation market, this figure still lags behind market demand and causes an inevitable bottleneck effect for the further development of China's civil aviation market. According to World Economic Forum's statistics, for all 133 countries surveyed, China's airport density is ranked 125 for 2009, while the quality of aviation infrastructure is ranked 74 for the 2008-2009 period (Schwab and Porter 2008).

¹ But Button and McDougall (2006) also indicate that some other commercialized ATM providers also had increased such costs of between 4% and 38%.

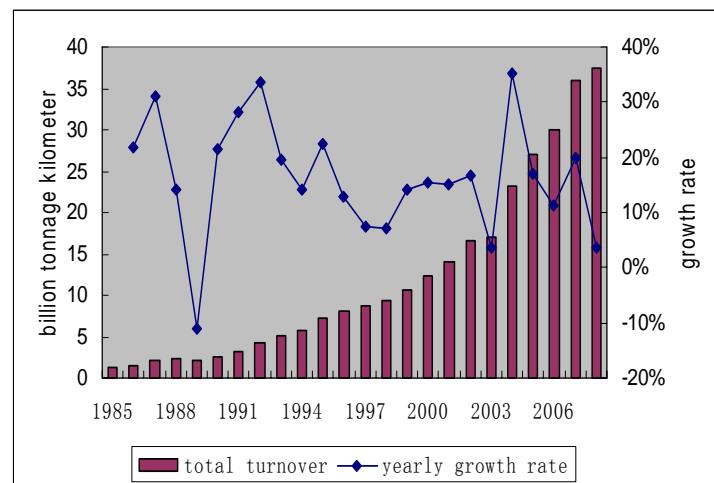


Figure 1 China's Air Transportation Market Development Pace (1985-2008)

Source: Civil Aviation Administration of China (2008).

The problem of China's ATM sector is even worse. Currently, the sole ATM service provider in China is the Air Traffic Management Bureau (ATMB) of the Civil Aviation Administration of China (CAAC). As a result of fast growing air service demand, ATMB finds that hub airports are reaching their capacity to handle aircraft landing and taking off. According to CAAC statistics, 17.52 – 26.11% flight delays in 2009 can be attributed to ATM reason (Civil Aviation Administration of China 2009)². Under these circumstance, CAAC has to set more restrictions so as to ease the serious flight delay problem. For example, CAAC set an official restriction for a daily slot ceiling in Beijing Capital Airport. The slot ceiling was originally set at 1,150 operations daily before July 2007, then it was reduced to 1,100 in July and was further reduced to 1,000 after late Oct 2007 (YU 2007). Besides, CAAC had to issue a ban that no new domestic airlines would be permitted to operate before 2010 so as to reduce pressure for airspace resources. But all these measures could only reduce the serious conditions for ATM temporarily, as it can not address the gap between fast growing demands and limited ATM capacity. New measures are badly needed so as to increase ATM efficiency and to accommodate fast growing air service demands.

As a result of above mentioned tensions, CAAC proposed that China's civil airports will amount to 190 by 2010 in the Outline of the Eleventh Five-Year (2006 – 2010) Plan for the civil aviation industry (Wang 2007). In 2008 CAAC published the master plan for China civil airports, where it was said that by 2020, China's civil airports would reach 244 (Civil Aviation Administration of China 2008). Also, government encourages the introduction of private entities in airport projects through a variety of regulation concerned. As to the ATM sector, CAAC tries to introduce significant institutional reform so as to increase the overall efficiency of China's ATM system. ATMB

² Those data are based on monthly statistics which mean that delays caused by ATM show significantly variation in different months of 2009

used to be a branch of CAAC which not only provides ATM service, but also plays a role as a regulator. In 2007, CAAC started the new round of reforms of the ATM segment. ATMB was transferred to a public institution which should operate in a business like manner. All the policy setting and regulatory function in the ATM area were transferred to the office of ATM affairs of CAAC. Another objective of the 2007 ATM reform is to integrate different levels of ATM services, as local ATM branches were used to be subordinates of local CAAC branches. After the 2007 reform, those local ATM branches were set as subordinates of ATMB and ATMB became an integrated ATM service provider across China. All those changes increase the autonomy of the ATM service provider and make it possible for ATMB to operate like a business organization.

V. CONSIDERATIONS FOR POSSIBLE PRIVATIZATION AS TO CHINA'S AVIATION INFRASTRUCTURES

It can be assumed that China's civil aviation market will still maintain its fast growth rate for quite a long time. But the capacity insufficiency of the aviation infrastructure will be the most obvious obstacle for this scenario. Obviously, that past government owned and operate model can no longer cope with developments in aviation market. Hence more commercialized and market oriented approaches are badly needed so as to improve operation performance of China's aviation infrastructure. That is the main reason why China wants to follow aviation infrastructure privatization approaches in developed economies. Also experience of China's past State Owned Enterprises (SOEs) reform can enlighten us in this regard, including following 3 recommendations for the Chinese aviation infrastructures:

A. Privatization should be considered as an intermediary measure for improving aviation infrastructure performance, rather than as an ultimate objective.

Ever since the introduction of economic reform in 1978, the reform of SOEs is a focus of policy makers. Partial privatization is also a widely applied approach, as many SOEs are changed into corporations and listed on stock exchanges (Aivazian, Ge et al. 2005; Chen, Firth et al. 2006). Some researches indicate that few years after a share issue privatization (SIP) program, there is a declining tendency in the profitability and performance for the privatized SOEs (Huang and Song 2005; Chen, Firth et al. 2006; Li, Moshirian et al. 2007). The privatization program itself is not necessary for the continuous improvement of operational performance. Especially some SOEs do not display well developed corporate governance structure during their privatization program. As for the aviation infrastructures, what really matters is to improve the performance of infrastructure services so as to accommodate soaring market demands. Privatization programs should not be treated as a one-off effort to generate capital needed for infrastructure investments.

B. Necessary regulation is needed so as to curb market power of monopolistic service providers

Even though new developments in economic theories indicate that potential entrants can place competitive pressure

upon the monopolistic firm in natural monopolies, one cannot deny the fact that firms concerned do enjoy certain extent of market power. Especially in China where generally there is only one airport in each city, the airport has significant market power. That is the reason why airports should be regulated even after the introduction of a privatization program. For ATM service, also it is true as more service providers are not feasible for this sector. After the introduction of a privatization program, it is still necessary to maintain a regulatory system so as to avoid market power abuses. Also, regulations are needed so as to maintain service quality and safety standards in airport and ATM services. But the regulatory framework should be adjusted so as to cope with privatization and interests of private partners. For example, currently CAAC sets the single fixed airport and ATM charges for all the aviation infrastructure services around China. Obviously, it does not consider the price change factor and cost differences among different places. It could be valuable to consider approaches taken by Western regulators like the CPI-X approach, which takes both the price change factor and operational efficiency increases into consideration.

C. Full privatization may not be appropriate at this stage, while partial privatization arrangement like corporatization and/or commercialization can be very helpful to improve operational performance of service providers in China's aviation infrastructure.

The main problem of China's aviation infrastructure lies with the low operational performance, rather than with insufficient capital. Past government direct management and monopolistic market position made service providers lack business incentives to improve their economic efficiency and accommodate user demands. Privatization programs can help to establish an autonomous corporate governance structure and place much more market pressures upon top management of those service providers. But on the other hand, aviation infrastructure services are also important public goods which are crucial for the overall civil aviation industry. Up to now, there are only limited cases of full privatization of airports and no such case in ATM. Considering the corporate governance conditions of China's aviation infrastructure service providers, it is much more important to establish well governed corporate structures so as to make airport and/or ATM service providers operate more efficiently. Radical full privatization approach is not suitable for China's airport and ATM sector, not only because full privatization is likely to cause political tensions in China, but also because of the fact that it may be optimal for governments to carry out corporatization of airport and ATM providers before eventual privatization.

VI. CONCLUSION

It is of no doubt there are pressing needs for institution changes together with technological improvement programs in China's aviation infrastructure sector, so as to accommodate fast growing demand for air transport service. And new institutions should place more emphases on aviation infrastructure operation performance rather than simply accommodating financing requirements for infrastructure projects. On considering the insufficient public spending in

China's aviation infrastructure sector compared with actual investment required and dissatisfactory performance of airport and ATC operators, privatization may be an effective and efficient approach for the financing and operation of China aviation infrastructures. But based on the characteristics of aviation infrastructures, together with the past experiences in China's State Owned Enterprise reform, partial privatization is a much ideal option for China aviation infrastructure sector. At the same time, it is suggested by theories and past experience that aviation infrastructure privatization program should be placed under strict scrutiny which can help to avoid the abuse of operator's market force. Consequently, China's civil aviation regulator should not only increase the market influences upon aviation infrastructure by the introduction of partial privatization arrangements, but also increase regulation concerned so as to avoid the market force abuse of service operators.

REFERENCES

- [1] Airbus (2009). Global Market forecast: 2009-2028.
- [2] Aivazian, V. A., Y. Ge, et al. (2005). "Can corporatization improve the performance of state-owned enterprises even without privatization?" Journal of Corporate Finance 11(5): 791-808.
- [3] Button, K. and G. McDougall (2006). "Institutional and structure changes in air navigation service-providing organizations." Journal of Air Transport Management 12(5): 236-252.
- [4] Chen, G., M. Firth, et al. (2006). "Have China's enterprise reforms led to improved efficiency and profitability?" Emerging Markets Review 7(1): 82-109.
- [5] Chen, K., Y. Tan, et al. (2004). "Zi Ran Long Duan Hang Ye Fan Song Gui Zhi De Li Lun Fen Xi (Theoretical Analysis on the Deregulation on Natural Monopoly)." Journal of South China Agricultural University (Social Science Edition) 3(1): 53-59.
- [6] Civil Aviation Administration of China (2008). Cong Tong Ji Kan Ming Hang (Statistical Data on Civil Aviation of China). Beijing, China Civil Aviation Press.
- [7] Civil Aviation Administration of China. (2008). "Quan Guo Ming Yong Ji Chang Bu Ju Gui Hua (Master Planning of National Civil Airports)." from http://www.caac.gov.cn/I1/I2/200808/t20080819_18371.html.
- [8] Civil Aviation Administration of China. (2009). "Zhong Guo Ming Hang Zhu Yao Yun Shu Sheng Chan Zhi Biao Tong Ji (Main Operational Statistics of China Civil Aviation Industry)." from <http://www.caac.gov.cn/I1/>.
- [9] De Neufville, R. and A. R. Odoni (2003). Airport systems : planning design, and management. New York, McGraw-Hill.
- [10] Donohue, G. L. (1999). "Air traffic service privatization: will the USA join other Developed Nations?" Journal of Air Transport Management 5(2): 61-62.
- [11] Forsyth, P. (2008). Airport Policy in Australia and New Zealand: Privatization, Light-hand Regulation and Performance, pp65-99. Aviation infrastructure performance : a study in comparative political economy. C. Winston and G. d. Rus. Washington, D.C., Brookings Institution Press: vi, 237 p.
- [12] Gillen, D. and H.-M. Niemeier (2008). The European Union: Evolution of Privatization, Regulation, and Slot Reform, pp36-64. Aviation infrastructure performance : a study in comparative political economy. C. Winston and G. d. Rus. Washington, D.C., Brookings Institution Press: vi, 237 p.
- [13] Goodliffe, M. (2002). "The new UK model for air traffic services--a public private partnership under economic regulation." Journal of Air Transport Management 8(1): 13-18.
- [14] Graham, A. (2008). Airport Planning and Regulation in the United Kingdom, pp100-135. Aviation infrastructure performance : a study in comparative political economy. C. Winston and G. d. Rus. Washington, D.C., Brookings Institution Press: vi, 237 p.
- [15] Huang, G. and F. M. Song (2005). "The financial and operating performance of China's newly listed H-firms." Pacific-Basin Finance Journal 13(1): 53-80.
- [16] Li, D., F. Moshirian, et al. (2007). "Managerial ownership and firm performance: Evidence from China's privatizations." Research in International Business and Finance 21(3): 396-413.
- [17] Lipovich, G. A. (2008). "The privatization of Argentine airports." Journal of Air Transport Management 14(1): 8-15.
- [18] Schwab, K. and M. E. Porter (2008). The global competitiveness report. Geneva, World Economic Forum: v.
- [19] United States. Government Accountability Office (2005). Air traffic control : characteristics and performance of selected international air navigation service providers and lessons learned from their commercialization : report to Congressional Requesters. Washington, D.C., GAO.
- [20] Vasigh, B., T. Tacker, et al. (2008). Introduction to air transport economics : from theory to applications. Aldershot, England ; Burlington, VT, Ashgate Pub.
- [21] Vickers, J. (2008). "privatization." The New Palgrave Dictionary of Economics, from <http://www.dictionaryofeconomics.com/article?id=pde2008_P000193>
- [22] Wang, Z. Q. (2007). "'Shi Yi Wu Gui Hua': Zhong Guo Ming Hang Ye De Fa Zhan Zhan Wang ('The Eleventh Five Year Planning': Developmental Vision for China's Civil Aviation Industry)." from http://www.caac.gov.cn/D1/MHLT/07MHLT/07YJZY/200805/t20080516_14609.html.
- [23] World Bank. (2009). "Subsector data for Transport: Airports." from http://ppi.worldbank.org/explore/ppi_exploreSubSector.aspx?SubSectorID=5.
- [24] YU, D. (2007). Kong Zhong Da Yong Du ("airfield congestion"). Cai Jing("Finance and Economy"). Beijing, 17.
- [25] Zheng, X. W., W. Lu, et al. (2009). Ming Hang Gai Ge Kai Fang 30 Nian Dui Jing Ji She Hui Gong Xian Yin Xiang Yian Jiu (Study On Social Economic Impacts Bought By 30 Years Economic Reform In China's Civil Aviation Industry).

Capturing the Impact of Fuel Price on Jet Aircraft Operating Costs with Engineering and Econometric Models

Megan Smirti Ryerson, Mark Hansen

University of California, Berkeley

National Center of Excellence for Aviation Operations Research

Berkeley, CA, USA

msmirti@berkeley.edu

Abstract— Challenges in forecasting fleet development and deployment are in part due to fuel price uncertainty. To address this issue, a recent study developed an aircraft-specific Leontief technology operating cost model (LM) to compare aircraft costs under fuel price uncertainty. This model considers individual aircraft types to be Leontief technologies, such that the key drivers of cost must be used in fixed quantities. While asserted in the literature that models in this form can more accurately predict operating costs, the Leontief specification precludes a precise examination of how aircraft size will change due to economic forces. To this end, an econometric operating cost model (EM) is developed. The translog functional form is used to capture the effect of the key drivers of cost on jet operating costs and also allow for substitution between inputs. A comparison of the LM and EM shows that the Leontief technology assumption limits the LM to capturing operating costs in only a snapshot in time, while the EM captures the input substitution that occurs with factor price changes. The conclusion that the EM has strong predictive potential encourages a strengthening of the model towards capturing costs related to passenger preferences. This study takes a total logistics cost approach (TLC) and considers passenger value of frequency along with operating cost to be the total cost per operation. The cost-minimizing seat size is smaller and more reflective of existing conditions under TLC compared with operating cost alone, yet the difference diminishes as fuel price increases. This study highlights the predictive potential of econometric cost models and also the importance of considering passenger preferences in predicting future aircraft economics.

Keywords—*Jet Aircraft; Operating Cost; Aircraft Size; Logistics Cost; Fuel Price; Leontief Technology; Econometric Model*

I. INTRODUCTION

Challenges in forecasting fleet development and deployment are in part due to fuel price uncertainty. Fuel price uncertainty is due to fuel and energy price fluctuations and a growing awareness of the environmental externalities related to transportation activities, particularly as they relate to climate change [1]. The impact of fuel price uncertainty is

The authors would like to thank the University of California Transportation Center for funding support.

evident in conflicting future fleet forecasts. The Boeing Current Market Outlook predicts the percent of regional jets in service will drop by 10 percent in 2028. This prediction is mainly due to predicted surges in the price of fuel as regional jets have lower fuel economy per seat than larger jets [2]. An increase in single and twin aisle aircraft is predicted over the largest jets because of their ability to balance operating costs with passenger preferences. In contrast, a fleet forecast performed by MITRE predicts a large increase in the percent of regional jets, in part due to surging passenger demand, and an increase in the largest aircraft due to cost savings potential [3]. The conflicting forecasts showcase the challenge of predicting how future fuel prices will affect fleet in the aviation system and also the importance of considering passenger demand and preferences in the forecasts. As airlines are considering new fleets and manufacturers are looking to meet future demands, research on the relationship between aircraft size and fuel price and the influence of passenger preferences on aircraft comparative costs can assist both parties in determining the aircraft type to best meet future cost pressures.

This study will 1. Investigate the potential of two operating cost models to capture the effect of fuel prices on aircraft economics and 2. Develop a Total Logistics Cost model by incorporating passenger preference cost and operating cost. The first cost model presented is an econometric operating cost model (hereafter, EM), in that it uses econometric methods to model operating costs based on airline-aircraft operating cost data. This model allows for detailed analysis on the interactions between the key drivers of cost and also allows for operating cost predictions over a range of fuel prices. However, such a process is data intensive, and the resulting model is cumbersome due to a long variable list. To this end, the EM developed in this study is compared with operating cost models which consider aircraft to be Leontief technologies (hereafter, LM) recently developed by Smirti and Hansen [4] to study aircraft comparative costs under fuel price uncertainty. The LM sums the key drivers of cost and allows for operating cost calculation and prediction with limited input needs. It is

asserted in [5] that models developed in this manner, termed engineering models, can lead to a more accurate cost functions; this study will explore this assertion by highlighting the unique contributions of econometric models and examining the relationship between the LM and EM estimates.

The comparison sheds light on the ability of EM to capture input substitution and therefore more accurately reflect operating costs and minimum-cost aircraft size under fuel price uncertainty. Therefore, this study also looks to strengthen the predictive power of EM by considering a total logistics cost function which sums operating and passenger costs related to service frequency. Passenger costs are captured through a passenger schedule delay function. The seat capacity which minimizes operating cost alone and the total logistics cost is determined for a range of fuel prices and distances traveled, to identify the aircraft types which provide the lowest costs for a range of future fuel and passenger preference scenarios.

The engineering model developed in [4] follows a long line of established aircraft engineering cost model literature [6-8]. Reference [4], however, departs from the literature in that it takes a total logistic cost approach and develops cost models for three representative aircraft using US DOT Form 41 data: a narrow body Boeing 737-400, an Embraer 145 regional jet, and an ATR 72-200 turboprop. Fleets of each vehicle category are compared for operating cost alone and total logistics cost over a range of fuel prices and distances, and the minimum cost fleet mix is determined. A limitation is the consideration of aircraft size as inelastic; as there are currently a wide range of aircraft sizes on the market, it is possible to consider aircraft size to be elastic.¹ In an attempt to generalize engineering aircraft cost models that are not specific to an aircraft type, Swan and Adler [5] develop two jet aircraft operating cost models using Boeing and Airbus aircraft data only: one for single aisle aircraft and one for double aisle aircraft. Limiting the data source to the two airframe manufacturers implicitly limits the aircraft types considered to mid-size and large aircraft. Furthermore, as the model is based on aircraft size and distance traveled, the model is not able to capture cost changes due to economic forces such as fuel price fluctuations. Additional studies considering cost economics of aircraft size related to stage length using engineering cost models prior to 1999 are well discussed in Wei and Hansen [10].

Reference [10] develops an econometric operating cost model for jet aircraft with elastic aircraft size at the aircraft-airline level in a departure from the literature discussed to this point. The model includes fuel price as a variable in an econometric operating cost model, yet it is not a key variable of interest. Reference [10] find that aircraft economies of scale exist, yet attenuate at longer stage lengths. The variables of interest are restricted to those that help investigate

¹ It is important to note that this was not always the case. In a 1986 article, Viton [9] expresses an interest in modeling costs with aircraft size as a continuous variable yet cites the limited aircraft sizes available during the study period as reason to perform an aircraft specific analysis.

economies of aircraft size – seat size and average stage length – and aircraft types that were commonly used in the study period of 1987-1998. The importance of considering a total logistic cost function with passenger and operating cost rather than individual cost components is demonstrated by comparing [10] and [11]. Using a nested logit model, [11] finds that an airline's market share experiences greater increases from increasing vehicle frequency rather than aircraft size. These findings point to the importance of balancing airline operating cost and passenger preference costs when choosing fleet mix and determining flight schedules. Beyond aviation, total cost studies considering a combination of operating passenger, and infrastructure costs have a long history in urban transportation [12].

The remainder of this paper is organized as follows: The following section reviews the data collected for the development of the EM and the modeling approach. Coefficient estimates are presented and interpreted based on the objective of the study. The EM and LM are then used to calculate operating costs for a range of aircraft types and fuel prices, and the results compared. Based on the strength of the EM, the model is used to predict aircraft operating costs over future fuel prices to determine the seat capacity that minimizes costs. Finally, a generalized cost function that sums operating and passenger value of service frequency is developed and used to perform similar predictions.

II. ECONOMETRIC OPERATING COST MODEL

A. Data Description

There are multiple variables which influence operating cost and over which an airline can assert control. Reference [10] includes aircraft size, labor and fuel factor prices, and average distance traveled as such variables. The model developed in this study extends these variables to others that influence operating cost: average aircraft age, technology age, and utilization. To develop the operating cost model, data from the US Department of Transportation (DOT) Form 41 is collected. Form 41 provides quarterly cost data and operating statistics broken down per airline and per aircraft type. Data was collected for all quarters between the years 1996 to 2006, inclusive. Data from which factor prices are derived and the independent variable, Direct Operating Cost, were collected from Form 41 Schedule P-5.2. This variable is termed Operating Cost per Departure (OCD). Ownership costs related to depreciation and rentals were eliminated from this total to capture operational costs only. The data collected to develop factor prices includes expenditures on Aircraft Fuels and Pilots and Copilots Salaries. Aircraft operating statistics were collected from Form 41 Schedule P05B.² These statistics, collected for scheduled and non-scheduled service,

² It is important to note that aircraft fuels is the actual cost of the fuel, without fuel taxes, any additional costs for the act of fueling the aircraft, or other charges. It is not the total cost related to fuel consumption, but rather the actual cost of fuel. The fuel tax exclusion has little impact as the tax on commercial aviation fuel was constant and minimal through the study period at \$0.044/gallon.

include gallons of fuel used; available seat miles; revenue aircraft miles, departures performed; and block hours, or the sum of hours an aircraft spends from gate to gate. From these prices and statistics, the unit price of fuel (UPF), the unit price of labor (PIL), average stage length (ASL), and aircraft seat capacity (Seat) are derived.

As the variable Seat is a key component of the study, a more detailed description of the derivation is presented. Many airlines operate identical aircraft types with different seat capacities determined by their business models. For example, a legacy carrier looking to lure business passengers may operate an aircraft with fewer seats and more differentiated service classes, while a low cost carrier may use a one-class configuration. To exclude any cost impacts to operating different configurations of the same aircraft, each aircraft type is assigned the weighted average seat size for that aircraft type. The resulting seats range from 49 to 360 seats for twenty three unique aircraft types (Table I).³

TABLE I. AIRCRAFT MODELS USED IN OPERATING COST ANALYSIS

Year of Introduction	Aircraft Model	Seats
1992	Canadair RJ-200/ER/-440	49
2001	Canadair RJ-700	68
2002	Embraer EMB-170	72
1982	BAE-146-200	88
1988	BAE-146-300	91
2004	Embraer EMB-190	100
1997	Boeing B-717-200	111
1990	Boeing B-737-500	113
2003	Airbus A318	114
1996	Airbus A319	123
1998	Boeing B-737-700/700LR	128
1988	Boeing B-737-400	143
1988	Airbus A320-100/200	148
1998	Boeing B-737-800	150
2001	Boeing 737-900	169
1996	Airbus A321	170
1982	Boeing B-767-200/ER	178
1983	Boeing B-757-200	184
1998	Boeing B-757-300	222
1986	Boeing B-767-300/ER	231
1995	Boeing 777-200/20LR/233LR	282
1997	Boeing B-767-400	286
1989	Boeing B-747-400	360

Data on aircraft age and utilization is collected from Form 41, Schedule B-43, which includes the total number of each aircraft model in service per airline and the year the airline began to operate them. The aircraft utilization (UTIL) variable, the block hours per quarter operated for each airline-aircraft pair, was derived from these statistics, as well as the average length of time an airline operates a particular aircraft type (AvgAge). Collected from publicly available sources was the first year of entry in service across domestic airlines for a specific aircraft type; this data was used to calculate the

³ As the data is restricted to US-based carriers, the aircraft in the dataset are those operated by US carriers.

technology age (TechAge) of the aircraft, or years that elapsed in between 2006 and the first year of aircraft service.

To capture the materials price, the Producer Price Index is collected from the Bureau of Labor Statistics; a similar method is employed in the work of [13] as well as [14] to develop airline cost functions. Instead of converting each year of data into constant dollars, this study follows [14] and uses the Producer Price Index as both a proxy for materials cost and also a gauge of changes in the economy and inflation. Similarly, a time trend variable is included to capture changes in operating cost over time.

To determine any data reporting inconsistencies, the data was cleaned with assistance from Database Products, the distributor of Form 41 data. The variables derived from the data sources are presented in Table II.

TABLE II. OPERATING COST MODEL VARIABLE DESCRIPTION

Variable Code	Variable Description (Units)
Dependent Variable	
OCD	Total aircraft operating expenses per departure (\$)
Independent Variables	
t	Time trend variable 19961: t=1...20064: t=44
Seat	Average seats per departure (Seats)
Util	Block hours in year-quarter t (Utilization metric) (Hours)
ASL	Average stage length (Miles)
Pil	Pilot salaries per block hour (\$)
PPI	Producer price index (Proxy for materials price and service)
UPF	Unit price of fuel (\$/Gallon)
AvgAge	Average years of aircraft operation by airline (Years)
TechAge	Aircraft technology age (Years)

The twenty six airlines (legacy, regional, and low cost) present in this study are shown in Table III.

TABLE III. AIRLINES USED IN OPERATING COST ANALYSIS

Airlines	
American	AirTran
Alaska	JetBlue
Continental	Midwest
Delta	Independence Air
Northwest	Trans World
United	Air Wisconsin
USAir	Atlantic Southeast
Southwest	Comair
America West	Horizon
National	Skywest
ATA	Hawaiian
Pinnacle	Aloha
Frontier	Spirit

B. Econometric Operating Cost Model Specification and Estimation Results

The model specification used is a demeaned translog model to estimate the operating cost per departure (OCD) (1). The translog model is widely used in cost modeling (for example, [10, 13, 14]); as a second order Taylor series expansion about the mean, it is able to approximate many different model specifications. The variables in the model are

defined by two indices that are the unique identifier of one observation: k indicates a unique airline code and aircraft type combination and q indicates year and quarter.

There are four groups of independent variables in the model. The first, α , is a time-invariant and aircraft airline group-invariant constant. The second, τt , is the time trend variable (t) and the coefficient to be estimated (τ). The third, A_k , are the airline-aircraft fixed effects, which capture the unobserved airline-aircraft effect. The variables X_{kq}^j represent the value of independent variable j for a given (k , q) combination (where j and ℓ are indices representing the $N=8$ independent variables that vary with a particular k). Independent variables $j = 1, 2, \dots, 6$ are transformed with the natural logarithm (Seat, Util, ASL, Pil, PPI, UPF) and independent variables $j = 8, 9$ not transformed with the natural logarithm (AvgAge, TechAge). Parameters ω_j and δ_{jl} are to be estimated.

The model is demeaned such that the dependent variable and the independent variables which vary across a given k are estimated about their mean values. This enables straightforward interpretations of the results: the average effect of each independent variable j is immediately evident from each parameter estimate ω_j .

$$\ln OCD_{kq} - \bar{\ln OCD}_{kq} = \alpha + \tau t + A_k + \sum_{j=1}^N \omega_j [X_{kq}^j - \bar{X}^j] + \sum_{j=1}^N \sum_{l \neq j} \delta_{jl} [X_{kq}^j - \bar{X}^j] [X_{kq}^l - \bar{X}^l] + \varepsilon_{kq} \quad (1)$$

This is a panel data set as there are k airline-aircraft groups over a set of year-quarters q . Because the elements of k are not constant across years, the panel is unbalanced. To estimate the model, a fixed effects mean-difference model is used, where the fixed effects are captured by A_k . Each observation in a particular group m (where $m \in k$) is estimated about the mean of group m . This method ensures consistent estimates of ω_j and δ_{jl} , however, the method precludes estimation of the time-invariant regressors A_k . As the data is over 44 time periods, the estimation method also corrects for autocorrelation present across airline-aircraft pairs. Finally, an examination of residuals shows heteroskedasticity across groups, and therefore generalized least squares with heteroskedastic-robust standard errors estimation is used ensuring consistent standard errors.

Table IV contains estimation results for the EM. The coefficient estimates generally have the expected signs and most are significant at the five or one percent level. The evaluation of operating cost economies of aircraft size (represented by the variable Seat) and fuel price (represented by the variable UPF) begins with the first order coefficient on aircraft size, .44. This implies operating cost economies of aircraft size; a one percent increase in aircraft size would increase operating cost by .44 percent. The second order term of Seat is positive (.27) and implies that aircraft economies of scale attenuate for aircraft sizes larger than the average size. There are economies of fuel price found, and the second order effects show that as fuel prices deviate positively from the

mean these cost economies of fuel price decrease. Finally, the interaction term between fuel price and aircraft size, 0.085, shows that as fuel prices increase, economies of scale due to aircraft size diminish slightly. In sum, economies of scale attenuate at larger aircraft sizes and at higher fuel prices, which confirms the assertion that “increases in fuel efficiency are harder to achieve in a larger plane” [15].

TABLE IV. JET AIRCRAFT EMPIRICAL RESULTS

Variable	Parameter Estimate	Standard Error
Constant	-0.129***	0.025
t	0.002***	0.001
Seat	0.436***	0.062
ASL	0.775***	0.041
Pil	0.346***	0.024
UPF	0.364***	0.026
Util	-0.056***	0.025
PPI	0.036	0.163
AvgAge ⁺	0.037***	0.004
TechAge ⁺	0.009***	0.002
Seat*Seat	0.273***	0.05
ASL*ASL	0.131***	0.013
Pil*Pil	0.045***	0.003
Util*Util	-0.020***	0.006
UPF*UPF	0.161***	0.026
AvgAge ⁺ *AvgAge ⁺	$5.7*10^{-4}$ *	$3.0*10^{-4}$
TechAge ⁺ *TechAge ⁺	$4.5*10^{-4}$ ***	$2.0*10^{-4}$
Seat*ASL	-0.187***	0.062
Seat*Pil	-0.129***	0.033
Seat*Util	-0.0247	0.0285
Seat *PPI	0.2845	0.2164
Seat*UPF	0.085***	0.037
Seat*AvgAge ⁺	-0.027***	0.007
ASL*UPF	-0.005	0.025
ASL*Pil	0.0117	0.0212
ASL*Util	-0.0436***	0.0179
ASL*PPI	0.0356	0.1492
ASL*AvgAge ⁺	0.0015	0.0039
Pil*Util	-0.0190	0.0127
Pil*PPI	-0.0215	0.0696
UPF*Pil	-0.103***	0.021
Pil*AvgAge ⁺	0.0084***	0.0032
PPI*AvgAge ⁺	0.0313*	0.0170
AvgAge ⁺ *UPF	-0.014***	0.003
UPF*PPI	-0.565***	0.185
Util*UPF	-0.021	0.019
Util*PPI	0.1700***	0.0748
Util *AvgAge ⁺	-0.0077***	0.0030
TechAge ⁺ *Seat	-0.0004	0.0049
TechAge ⁺ *ASL	-0.0011	0.0037
TechAge ⁺ *Pil	-0.0011	0.0022
TechAge ⁺ *UPF	$3.54*10^{-3}$	$2.27*10^{-3}$
TechAge ⁺ *Util	-0.0057***	0.0021
TechAge ⁺ *PPI	0.0166	0.0138
TechAge ⁺ AvgAge ⁺	-0.002***	$4.5*10^{-4}$
N obs		1657
N groups		66

***Variables are significant at the 1% level

**Variables are significant at the 5% level

*Variables are significant at the 10% level

⁺Variables are not natural log

The negative sign on the interaction term between distance traveled (represented by the variable ASL) and fuel price confirms that there are more scale economies over longer distances due to fuel consumption. As the cruise phase is the most efficient from a fuel consumption perspective, this is the expected result. This finding is further reinforced by the interaction term between aircraft size (Seat) and distance traveled (ASL), which shows that at longer distances traveled there are strong economies of operating cost due to aircraft size.

While previous studies have excluded the technology age (TechAge) and the average age (AvgAge) variables, the model estimates show that the inclusion of these variables is warranted by their significant effect. The negative interaction term between average age and fuel price is unexpected, and could be explained by airline comfort with aircraft. As an airline learns how to operate an aircraft with experience, it learns the optimal fuel level and optimal flying speeds and altitudes. Such benefits are found by Southwest Airlines and their one aircraft type fleet [16]. The interaction of technology age and fuel price is not statistically significant, yet the sign of the coefficient tells us that as an aircraft ages it is more impacted by fuel prices. The interaction between aircraft size and average aircraft age shows that smaller aircraft show the signs of age more quickly, as a larger aircraft has more cost economies due to size than a smaller aircraft of the same age.

III. COMPARISON OF OPERATING COST MODEL RESULTS UNDER FUEL PRICE UNCERTAINTY

This section will use the EM to calculate operating costs for a range of inputs and compare these results with the LM.

A. Econometric Operating Cost Model Analysis

Using the coefficient estimation results presented in Table IV and other assumed inputs, operating cost per seat mile over a range of stage lengths for pairs of aircraft sizes and prices of fuel is calculated. The operating cost calculation is done by estimating the cost functions at certain specified values. The results presented will be parametric over fuel price and stage length; combinations of these two variables will be specified inputs. The PPI and time trend variable will be set at the 2006 value, and the values reported will be in 2006 dollars. For labor costs, a simple univariate linear model that relates the dependent variable, the unit price of labor, to the independent variable, seats is developed. The following is the resulting equation, with all coefficients significant at the 5 percent level.

$$Pil = 140.1 + 1.78 * Seat \quad (2)$$

For the remaining variables, the average factor prices and aircraft operating statistics for Delta Airlines will be used.

Fig. 1 presents the results.⁴ There is a unique minimum operating cost per seat mile for each aircraft size, dependent on the average stage length flown and the fuel price. For constant fuel price, as the distance flown increases, the aircraft

⁴ Fig. 1 presents four representative stage lengths for ease of presentation.

size which minimizes operating cost per seat mile increases. This finding is consistent with the negative interaction term between seats and average stage length. For a constant distance flown, as fuel price increases, the aircraft size which minimizes operating cost per seat mile increases; while the interaction term between seats and fuel price is positive, the interaction between labor and fuel price is negative.

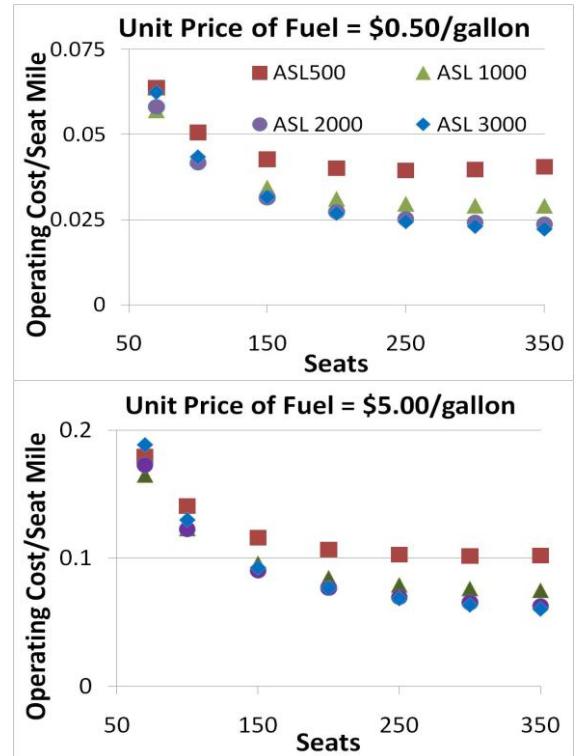


Figure 1. Operating cost per seat mile vs. seats for representative fuel prices.

B. Leontief Technology Operating Cost Model Comparison

This section will investigate the difference in predicted values between the LM and EM developed in section II of this study. The LM was developed by Smirti and Hansen [4] using average values from the same data set used in the current study, but for the year 2007. In [4], three specific aircraft types are chosen for cost calculation, two of which are jet aircraft: an ERJ 145 regional jet (50 seats) and a Boeing 737-400 narrow body (141 seats). The key drivers of cost including fuel costs, labor costs, and maintenance costs are summed based on statistical relationships between fuel burn and distance traveled and travel time and distance traveled.⁵ The values presented in [4] are reported in Table V. Using the same methodology as in [4], the cost coefficients for a mid-sized aircraft, the narrow body Boeing 757-200 are determined so the comparison can cover aircraft with ranges up to 3000 miles. The values calculated for the Boeing 757-200 are reported in Table V.

⁵ It should be noted that [4] also includes airport charges as part of the operating costs; these are eliminated for this analysis because they are not part of the direct operating costs.

To perform the comparison of LM and EM results, three key inputs are necessary: fuel price, distance traveled, and seat size. The seat size input is only necessary in the EM: the three set seat sizes for the aircraft in the LM are used. Three representative fuel prices: \$0.50/gallon, \$3.00/gallon, and \$5.00/gallon are used and stage lengths between 100 and 3000 miles are used as the additional inputs.⁶

TABLE V. OPERATING COST PER DEPARTURE EQUATIONS

Aircraft Category	Coefficient Value			
	Fuel Price (f)	Distance*Fuel Price (d*f)	Distance (d)	Fixed
B757-200	5.1×10^2	2.0	2.5	9.4×10^2
B737-400	2.7×10^2	2.1	2.6	8.8×10^2
ERJ 145	1.9×10^2	1.9	1.2	4.8×10^2

The LM estimates, developed using the inputs and the values in Table V, are compared with the EM estimates calculated in Section III(A). For comparison, the values are plotted against each other for the three aircraft types in Fig. 2.

Fig. 2 shows a relatively linear relationship along the 45-degree equality line between the LM and EM for the three aircraft at each fuel price. However, there is under prediction by the LM present at low fuel prices and over prediction by the LM for high fuel prices. This is due to the technology assumptions behind the EM and LM. The LM considers aircraft to be a Leontief technology, in that all inputs must be used in fixed proportions. The EM model allows substitution between inputs when factor prices change.

The LM was developed at a time when the operators of a 737-400 were paying an average of \$2.01/gallon; the operators of a 757-200 were paying an average of \$1.99/gallon, and operators of the ERJ 145 were paying an average of \$1.22/gallon. The average fuel price for Delta Airlines in 2006, the year of the projection data for the EM, is \$2.08 per gallon of jet fuel. It therefore follows that when the EM and LM are estimated at fuel prices close to this \$2.00/gallon average, the EM predictions and the LM predictions will be close. For fuel prices above this average, the LM should have higher estimates than the EM. This is because the EM allows for input substitution: as fuel prices increase, airlines will take steps to use fuel more efficiently by leveraging other inputs, a technical infeasibility of the LM. This hypothesis is confirmed in Fig. 2.

IV. OPERATING COST AND TOTAL LOGISTICS COST COMPARISON

The EM proves to be a useful predictor of operating cost, as it is able to capture costs in the current and future environment. To this end, the EM is improved by adding to it a passenger cost component. Schedule delay, or the concept

⁶ As the ERJ 145 has a range of 1,550 miles, the operating cost estimation is not performed for distances further than 2000 miles. The B737-400 has a range of 2,255 miles, and the cost estimation is not performed for distances further than 2,500 miles.

that passengers place a value on the difference between desired arrival time and actual arrival time is well known to airlines and manufacturers. This generalized cost model incorporates schedule delay and is termed the Total Logistics Cost (TLC) function. In this section, this model is used to compare the aircraft sizes that optimize operating cost alone and TLC for a range of fuel prices.

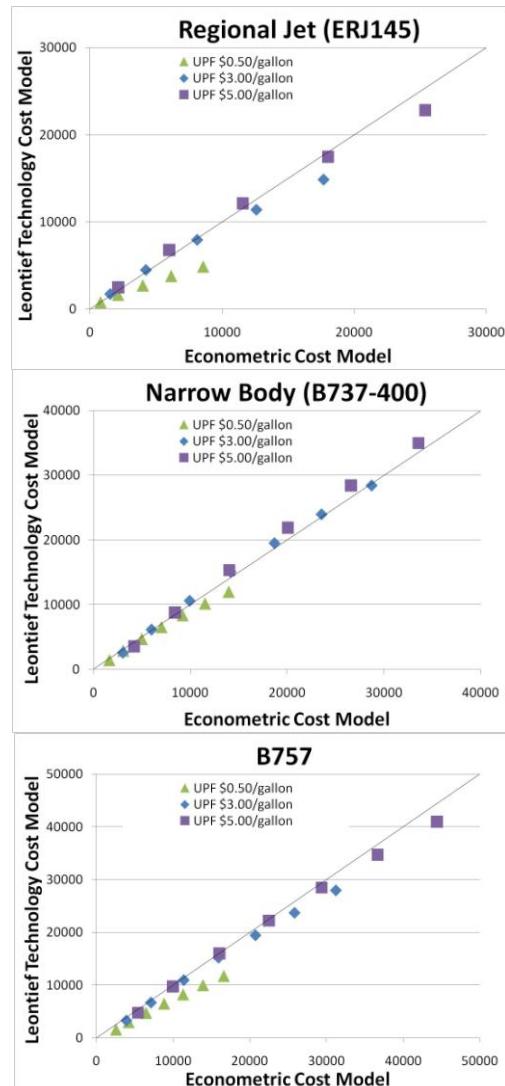


Figure 2. Leontief technology vs. econometric operating cost model results.

A. Total Logistics Cost Analysis

As the consideration of operating cost alone does not capture the entire motivation behind fleet adoption and utilization decisions, this study develops a generalized cost function including operating cost and passenger schedule delay cost. To capture schedule delay costs, two relationships must be determined, one between vehicle size and frequency and the other between frequency and schedule delay. Reference [17] develops a relationship for frequency and schedule delay based on flight frequency, which accounts for schedule peaking. Equation (4) shows the schedule delay

function $g(f_i)$ in hours based on a frequency function (3).⁷ The equation for flight frequency (f_i) is determined by the market density, or the passenger flow per day between a given origin and destination per day (q); the aircraft seat capacity (s); and the load factor, or the percent of seats occupied per departure (ℓ). The resulting schedule delay, $g(f_i)$, can be multiplied by the weighted average of schedule penalties for business and non business travelers, λ_{SD} (\$15.77/hour [18]). Delays in either direction (early or late) are considered equally onerous.

$$f_i = q / (\ell * s) \quad (3)$$

$$g(f_i) = 5.7 / f_i \quad (4)$$

B. Determining of Minimum Cost Seat Capacity for Operating Cost and Total Logistics Cost

1) Minimum Operating Cost Seat Capacity

To find the seat size which minimizes operating cost per seat mile, the operating cost function (1) is minimized for each stage length and fuel price combination. The results are shown in Table VI. For a constant stage length, the seat size which minimizes operating cost per seat mile increases with fuel price, yet at a decreasing rate. Certainly, as fuel price increases, the cost economies of aircraft size are stronger; while this is evident, the aircraft sizes for each stage length in Table VI are much larger than those seen today. As noted in the previous subsection, passenger preference for level of service, or schedule frequency, is an important component of airline decision of aircraft deployment.

TABLE VI. SEAT SIZE CORRESPONDING TO THE MINIMUM OPERATING COST PER SEAT MILE FOR A RANGE OF FUEL PRICES AND STAGE LENGTHS

ASL		UPF			
		0.5	1	3	5
	100	143	148	157	161
500	255	271	297	310	
1000	332	354	393	413	
1500	387	414	464	489	
2000	431	476	526	552	
2500	470	506	572	606	
3000	504	544	617	654	

2) Minimum-Total Logistics Cost Seat Capacity

The aircraft seat size that minimizes the generalized cost function, the TLC, over a range of fuel prices and stage lengths is the seat size that minimizes the operating cost function plus $g(f_i) * \lambda_{SD}$. Table VII shows the aircraft seat size that minimizes TLC over a range of fuel prices and stage lengths. Three representative market densities are chosen: a relatively low market density of 250 passengers/day, a medium market density of 750 passengers/day, and a high market density of 3000 passengers/day. The load factor is set to one. The values in the tables on the left side are the

⁷ It should be noted that other representations of schedule delay functions exist. Using pre-regulation data, Ref. [6] discusses a similar relationship which is less sensitive to high and low market densities. However little discrepancy between these two representations was found in [4].

solutions to the minimization of the TLC and the values in the table on the right are the percent difference between seat capacities before and after the inclusion of schedule delay.

TABLE VII. SEAT SIZE CORRESPONDING TO THE MINIMUM TOTAL LOGISTICS COST COST PER SEAT MILE FOR A RANGE OF FUEL PRICES, STAGE LENGTHS, AND MARKET DENSITIES

Market Density = 250 Passengers per Day						
ASL		UPF				
		0.5	1	3	5	
100	40	43	52	59		-72%
500	69	74	91	105		-73%
1000	92	99	124	144		-72%
1500	112	120	152	176		-71%
2000	129	139	176	205		-70%
2500	144	157	198	231		-69%
3000	159	173	219	256		-68%

Market Density = 750 Passengers per Day						
ASL		UPF				
		0.5	1	3	5	
100	61	65	78	88		-57%
500	105	113	140	160		-59%
1000	141	152	190	218		-58%
1500	170	184	231	266		-56%
2000	180	213	267	307		-58%
2500	228	239	300	345		-51%
3000	250	263	331	381		-50%

Market Density = 3000 Passengers per Day						
ASL		UPF				
		0.5	1	3	5	
100	96	102	117	127		-33%
500	166	179	214	237		-35%
1000	180	239	288	320		-46%
1500	282	287	347	385		-27%
2000	320	328	397	441		-26%
2500	354	365	441	490		-25%
3000	385	398	482	535		-24%

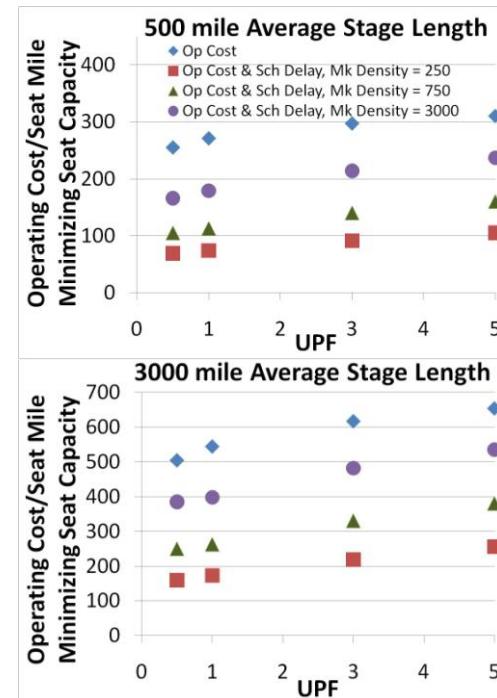


Figure 3. Seat size corresponding to minimum TLC per seat mile for a range of fuel prices, stage lengths, and market densities.

For all three market densities, the seat capacity that minimizes the TLC is reflective existing aircraft fleets. This is clear from Fig. 3, which shows the seat capacity which minimizes four scenarios of cost (operating cost alone, and the TLC function for the three market densities). Holding fuel price and stage length constant, as market density increases, the aircraft size which minimizes TLC increases. Higher demand necessitates larger aircraft sizes (Fig. 3); we would also expect this trend to appear if we were to decrease λ_{SD} (6).

By comparing the upper and lower panels of Fig. 3, it can be seen that an increase in stage length leads to an increase in seat size which minimizes cost per seat mile, holding fuel price and market density constant. Finally, across common market densities and stage lengths, as fuel price increases, the percent difference decreases. This is because in the generalized cost function, the operating cost becomes the dominant cost.

V. CONCLUSIONS

This study helps shed light on airline choice of aircraft size; as airlines are not looking to minimize operating cost alone but rather considering profit and market share, a strong weight is put on passenger preference when considering aircraft deployment. The difference observed in the minimum cost aircraft with the incorporation of passenger costs points to the importance of considering multiple costs when evaluating aircraft types. Results of this study show that the consideration of passenger preferences erodes as fuel price increase and that high fuel prices rationalize the use of larger aircraft in fleet composition despite higher passenger costs. Therefore, if fuel prices were to include other costs such as environmental taxes, the advantage of larger aircraft would be evident to airlines and airframe manufacturers.

This study also highlights the predictive potential of econometric operating cost models. The Leontief technology operating cost model has many strengths: transparency, few inputs, and the ability to provide predictions at a snapshot in time. The econometric model, in comparison, is shown to make predictions at a point in time and also capture how an airline might adapt to changes in factor prices. Both models play an important role in the aviation cost modeling space. However, this study shows the strengths of econometric cost models and their ability to provide consistent estimates and deep insight into current and future aircraft cost economics.

ACKNOWLEDGMENT

The authors would like to thank Lucretia Frederich of Database Products for assistance with formatting and cleaning the Form 41 Data. The authors would like to thank Bo Zou of the University of California, Berkeley for data analysis

assistance, Dr. Gautam Gupta of NASA for data collection assistance and comments on earlier paper drafts, and Amy Kim of the University of California, Berkeley for comments on an earlier paper draft.

REFERENCES

- [1] Smirti, M., and M. Hansen (2009): *Assessing the Role of Operating, Passenger, and Infrastructure Costs in Fleet Planning under Fuel Price Uncertainty*. [Online] Proceedings of the Eighth Annual FAA/EUROCONTROL Air Traffic Management Research and Development Seminar, Napa, CA, June 29-July 2, 2009 Available at: <http://www.atmseminar.org/8th-seminar-united-states-june-2009/papers/> [Accessed 30 July 2009].
- [2] Boeing Company. *Current Market Outlook 2009-2028*. [Online] Available at: http://active.boeing.com/commercial/forecast_data/index.cfm [Accessed July 13, 2009].
- [3] Hollinger, K.V. (2007): 'US Air Transport Fleet Forecast 2007 – 2035', MITRE Corporation.
- [4] Ryerson, M.S., Hansen, M. (2010): 'The potential of turboprops for reducing aviation fuel consumption', *Transportation Research Part D*, doi:10.1016/j.trd.2010.03.003.
- [5] Swan, W. M. and Adler, N. (2006): 'Aircraft Trip Cost Parameters: A Function of Stage Length and Seat Capacity,' *Transportation Research Part E*, 42, 105–115.
- [6] Douglas, G.W., and J.C. Miller III. (1975): *Economic Regulation of Domestic Air Transport: Theory and Policy*, The Brookings Institution, Washington DC, ch. 6.
- [7] Oster, C.O. Jr., and A. McKey (1984): 'Cost Structure and Short-Haul Air Service', in Meyer, J.R., C.O. Oster Jr. (eds.) *Deregulation and the New Airline Entrepreneurs*, MIT Press, Cambridge, MA.
- [8] Bailey, E.E., D.R. Graham, and D.P. Kaplan (1985): *Deregulating the Airlines*, The MIT Press, Cambridge, Massachusetts.
- [9] Viton, P. (1986): 'Air Deregulation Revisited: Choice of Aircraft Load Factors, and Marginal-Cost Fares for Domestic Air Travel', *Transportation Research Part A*, 2(5), 361-371.
- [10] Wei, W., and M. Hansen (2003): 'Cost Economics of Aircraft Size', *Journal of Transport Economics and Policy*, 37(2), 279–296.
- [11] Wei, W. and Hansen, M. (2005): 'Impact of aircraft size and seat availability on airlines' demand and market share in duopoly markets', *Transportation Research Part E*, 41(4), 315–327.
- [12] Meyer, M.D. and Miller E.J. (1984): Urban transportation planning: A decision-oriented approach. McGraw-Hill, New York, NY, pp. 171–307.
- [13] Caves, D.W., L.R. Christensen, and M.W. Tretheway (1984): 'Economies of Density Versus Economies of Scale: Why Trunk and Local Service Airlines Differ', *Rand Journal of Economics*, 15, 471-489.
- [14] Hansen, M. M., D. Gillen, and R. Djafarian-Tehrani (2001): 'Aviation Infrastructure Performance and Airline Cost: a Statistical Cost Estimation Approach', *Transportation Research Part E*, 37, 1-23.
- [15] Brueckner, J.K., and A. Zhang (2009): 'Airline Emission Charges: Effects on Airfares, Service Quality, and Aircraft Design', *CESifo Working Paper Series*, No. 2547.
- [16] Gittell, J.H. (2002): *The Southwest Airlines Way: Using the Power of Relationships to Achieve High Performance*, McGraw-Hill, New York.
- [17] Eriksen, S.E. (1978): Demand models for US domestic air passenger markets. Department of Aeronautics and Astronautics, MIT, Report No. FTL-R78-2, Cambridge, Massachusetts.
- [18] Adler T., C.S. Falzarano, and G. Spitz (2005): 'Modeling Service Trade-offs in Air Itinerary Choices', Presented at 84th Annual Meeting of the Transportation Research Board, Washington, D.C.

En route charges for ANSP revenue maximization

Alessia Violin and Martine Labb  

D  partement d'Informatique
Universit   Libre de Bruxelles
Boulevard du Triomphe, CP 210/01
B-1050 Bruxelles, Belgium
Email: aviolin@ulb.ac.be, mlabbe@ulb.ac.be

Lorenzo Castelli

Dipartimento di Elettrotecnica,
Elettronica e Informatica
Universit   degli Studi di Trieste
Via A. Valerio 10, 34127 Trieste, Italy
Email: castelli@units.it

Abstract—In Europe, all Air Navigation Service Providers (ANSPs) finance their activities by charging airlines using their airspace. These ‘en route charges’ usually account for a significant part of the cost of a flight, and they can therefore influence the route choice: airlines may decide to fly longer routes to avoid countries with higher charges. If ANSPs want to maximize their revenues, they must choose the optimal charge to impose on their airspace. We show that this optimal charge can be identified through a Network Pricing Problem (NPP) formulation in the form of Bilevel Programming where the leader (i.e. the ANSP) owns a set of arcs (the airways in its national airspace) and charges the commodities (i.e. the flights) passing through them. As the en route charges are proportional to a Unit Rate value fixed by the ANSP, we are able to apply a similar methodology as in the case of a single toll arc for the NPP. By exploiting the structure of the problem, we propose an exact algorithm to compute the optimal Unit Rate and apply it to a case study relying on real air traffic data and realistic flight cost figures.

Keywords - Network Pricing Problem, En route charges, Air Navigation Service Providers, Air Traffic Management

I. INTRODUCTION

In accordance with European Commission Regulation 1794/2006 laying down a common charging scheme for air navigation services, every European Air Navigation Service Providers (ANSP) finances its activities by charging airlines to use its airspace through the mechanism of ‘air navigation service charges’, which are charges levied on all flights passing through the ANSP’s airspace. These charges are composed of en route and terminal charges which are levied to finance costs for providing en route and terminal services, respectively. As en route charges linearly depend on a national Unit Rate which is fixed annually by the ANSP [1] and usually account for around 10-20% of the cost of a flight, the route choice can be influenced by them: airlines may decide to fly longer routes to avoid countries with high Unit Rates [2]. Currently, in most European states (except for the United Kingdom), the Unit Rate is set to allow the ANSP to completely recover all the costs it incurs to provide air navigation services, without making a profit. Over the next few years however many ANSPs, which nowadays are mostly not-for-profit corporations, are likely to move to more commercial approaches to the supply

of air navigation services [3]. In this scenario, ANSPs would instead aim to fix their Unit Rates so as to maximize their revenues.

In this paper we propose a Network Pricing Problem (NPP) formulation to identify this optimal Unit Rate value, in the form of Bilevel Programming (see [4]) where the leader (i.e. the ANSP) owns a set of arcs (the airways in its national airspace) and charges the commodities (i.e. flights) passing through them. Flights are assumed to have a rational behavior and look for the minimum cost path through the network. We prove that the NPP approach to fix the charge on a single toll arc (e.g. see [4]) can be extended to our case where the charge on each arc is proportional to a constant. In fact, as the Unit Rate is unique for each country and the charge to be paid on an arc linearly depends on it, the leader has to decide on this single value only.

Our findings show that flight travel choices do depend on the Unit Rate value set by the ANSP, and we also identify the revenue-maximizing Unit Rate value.

The paper is organized as follows: the following section will briefly introduce the Network Pricing Problem and give some references to studies on this topic, and then third section will describe the structure of en route charges in Europe. Section four will present our model along with the computational procedure proposed to solve it, and finally in section five we present some results from a preliminary case study. The last section will summarize our findings and present some discussions of them.

II. THE NETWORK PRICING PROBLEM

Consider a sequential game with two players, a leader L and a follower F . L plays first and decides his best strategy, taking into account the optimal strategy of F in reaction to his choice. F plays second, and so already knows L ’s choice of strategy when choosing his own. This is commonly known as a Stackelberg game [5] and has been widely studied in literature.

Bilevel programming (BP) provides an appropriate framework for modeling sequential games of this kind. A BP

problem is a hierarchical optimization problem in which the constraints are defined by a second optimization problem. This formulation was introduced by [6], and several studies have followed. By setting x as the decision vector of the leader and y as the decision vector of the follower, the general formulation is:

$$\begin{aligned} \max_{x,y} \quad & F(x,y) \\ y \in \arg \min_y \quad & f(x,y) \\ & g(x,y) \leq 0 \end{aligned}$$

This type of problem has been shown to be *NP*-hard even for linear objective functions or local optimality [7]–[9]. Several types of algorithms have been implemented in literature, and a literature review can be found in [9]–[11].

The Network Pricing Problem (NPP) is a type of Stackelberg game, which is based on a network, with an authority which owns a subset of arcs and imposes tolls on them, and users who travel on the network. The authority is the leader who wants to maximize his revenue, and network users are the followers who want to minimize their costs, and so will always travel on the minimum cost path.

The transportation network is defined as a set of nodes linked by a set of arcs. A commodity is a network user who travels from an origin to a destination and has some fixed cost parameters. To avoid a trivial solution, an assumption is made that for each commodity there exists a toll free path, which does not pass through any of the arcs owned by the authority. This condition avoids the possibility of the authority imposing an infinite toll on its arcs, which would lead to infinite revenues.

The NPP can therefore be modeled using bilevel programming. The bilinear/bilinear¹ bilevel Network Pricing Problem was first introduced by [4] for a multicommodity network. We adopt the notation used by [12]:

- $i \in \mathcal{N}$ nodes
- $a \in \mathcal{A} \cup \mathcal{B}$ arcs (\mathcal{A} is the set of toll arcs)
- $k \in \mathcal{K}$ commodities with demand η^k
- c_a travel cost of arc a , exclusive of toll
- $\forall k \in K: (o^k, d^k)$ origin/destination of commodity k
- t_a toll on arc $a \in \mathcal{A}$ (imposed by the authority)
- x_a^k flow of commodity k on arc a ($x_a^k = 1$ if commodity k travels on arc a , 0 otherwise)

The multicommodity NPP has been formulated as follows:

¹This means that the objective functions of both leader and follower are bilinear.

$$\begin{aligned} \max_{t,x} \quad & \sum_{k \in \mathcal{K}} \sum_{a \in \mathcal{A}} \eta^k t_a x_a^k \\ t_a \geq 0 \quad & \forall a \in \mathcal{A} \end{aligned} \tag{1}$$

$$x \in \arg \min_x \sum_{k \in \mathcal{K}} \left(\sum_{a \in \mathcal{A}} (c_a + t_a) x_a^k + \sum_{a \in \mathcal{B}} c_a x_a^k \right) \tag{2}$$

$$\begin{aligned} \sum_{a \in i^- \cap \mathcal{A}} x_a^k + \sum_{a \in i^- \cap \mathcal{B}} x_a^k - \sum_{a \in i^+ \cap \mathcal{A}} x_a^k \\ - \sum_{a \in i^+ \cap \mathcal{B}} x_a^k = \begin{cases} -1 & \text{if } i = o^k \\ 1 & \text{if } i = d^k \\ 0 & \text{otherwise} \end{cases} \\ \forall k \in \mathcal{K}, \forall i \in \mathcal{N} \end{aligned} \tag{3}$$

$$x_a^k \in \{0, 1\} \quad \forall k \in \mathcal{K}$$

where i^- and i^+ respectively denote the sets of arcs with i as tail or head.

In [4] the authors show that the lower level problem can be replaced by its primal dual constraints and primal dual optimality conditions, yielding a single-level problem. Many techniques have been applied to the NPP to obtain efficient algorithms and improved numerical results. For a deeper mathematical investigation and for a literature review of this problem, see [12], whose work concerns the particular case in which all toll arcs are connected and constitute a path, as occurs on motorways. Another interesting piece of work on this subject can be found in [13], which includes a large review of pricing in networks.

A. The case of a Single Toll Arc

As it will be useful later on in our description, we will now discuss the case of a Network Pricing Problem where the authority owns only one arc a . This is a relatively straightforward case, which can be solved using the parametric linear programming technique [4]. We define T as the tax value the leader can impose on arc a , and $\gamma_k(T)$ as the cost of the shortest path for the commodity k for a given value of T . We set the upper bound to the toll that can be imposed from the leader for commodity k as $\pi_k = \gamma_k(\infty) - \gamma_k(0)$. Then we sort all π_k quantities for all commodities in decreasing order. We assume that the order is $\pi_{k_1} \geq \pi_{k_2} \geq \dots \geq \pi_{k_{|\mathcal{K}|}}$. For any toll value T which is not equal to one of the values in this π_k sequence, we can increase the toll with $\epsilon > 0$ and achieve a higher revenue. Thus, every optimal value of T is equal to one of the π_k values. Moreover, for a toll value π_i ($i \in \{1 \dots |\mathcal{K}| \}$) only commodities $k \leq i$ (for which $\pi_k \geq \pi_i$) will choose the toll arc. The leader revenue function is:

$$\Pi(\pi_i) = \sum_{k \leq i} \pi_i \eta^k \tag{4}$$

where η^k is the demand for commodity k . The leader will choose the toll value that maximizes his revenue, so the

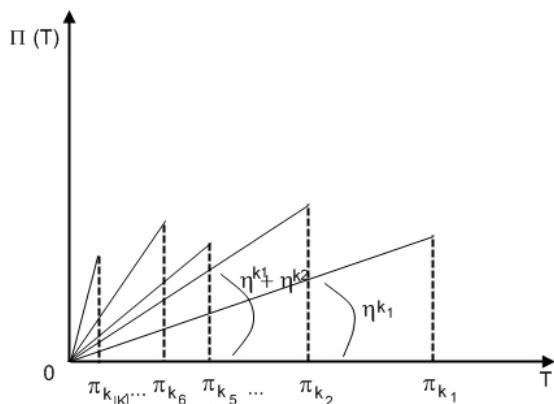


Figure 1. Leader revenue in the case of a Single Toll Arc

optimal solution will be:

$$T^* = \pi_{i^*}, \text{ such as } i^* = \arg \max_{i \in \{1, \dots, |K|\}} \Pi(\pi_i) \quad (5)$$

The leader revenue function is shown in the graph in Figure 1. It is a piecewise linear function, with discontinuities at π_i values. In each interval the function is described by a straight line which is linearly dependent on the cumulative demand of commodities which will choose the toll arc for that π_i value.

III. EN ROUTE CHARGES IN EUROPE

Although European Air Traffic Control is centrally coordinated, every country in Europe has an ANSP which manages flights within its national airspace. The air navigation service charges imposed by ANSPs to finance their activities are both a source of revenues for the ANSPs and costs for airspace users such as airlines.

For each flight, the en route charge is calculated using three basic elements [1]:

- Aircraft Weight Factor
- Distance Factor
- Unit Rate of en route charges (for each Charging Zone, i.e. each country)

The Weight Factor (expressed to two decimal places) is determined by dividing the maximum take-off weight (MTOW) of the aircraft (in metric tonnes, to one decimal place) by 50, and subsequently taking the square root of the result rounded to the second decimal, i.e. $w = \sqrt{MTOW/50}$.

The Distance Factor for each Charging Zone is obtained by taking the number of kilometers in the so-called ‘Great Circle Distance’², between either the aerodrome of departure or the entry point of the zone and either the aerodrome of arrival or the exit point of the zone, and dividing it by 100. This operation is repeated for each Charging Zone which the flight passes through. The entry and exit points are the points at

²The Great Circle Distance is the shortest distance between any two points on the surface of a sphere measured along a path on this surface (as opposed to going through the sphere’s interior).

which the lateral limits of the Charging Zone are crossed by the route described in the last plan filed.

The Unit Rate of en route charges is fixed by each ANSP and is the charge imposed on a flight per 100km flown within a given charging zone, and per 50 metric tonnes of aircraft weight. The Unit rates are applicable from 1st January of each year.

In literature, to the best of our knowledge, there are very few studies on air navigation charges in Europe. In [14] the congestion problem in European airspace is approached with the aim of applying a pricing solution. First an analysis of the formula used to calculate en route charges is performed and an explanation on why it is inefficient in preventing congestion is provided. Then a new formula is provided, with some congestion costs. Whilst the work is very interesting, it is not about the same type of analysis which we would like to develop on en route charges in Europe (how ANSPs should fix their Unit Rate value).

Another work on en route charges can be found in [15], where the authors provide a study on pricing schemes in the case of a unified upper airspace between certain countries. They propose some different scenarios of en route charges and analyze their impact on the actors involved (ANSPs and aircraft operators), but they do not propose a mathematical model to calculate an optimal charge.

In [16] the airport pricing models are analyzed and transposed for air carriers, whether they have market power or not. A pricing model for security charges on air travel is provided in [17]. Both of these works are interesting but they do not specifically deal with en route charges or with ANSPs’ behavior when fixing their Unit Rate values, which is the central topic of our study.

In this paper we would like to analyze the choice of the Unit Rate to fix every year as a pricing problem for European ANSPs, with a mathematical model able to determine the optimal value. To assess the validity of this approach, we will now look at how much influence en route charges have in affecting the cost of a given flight.

In [2] the authors provide an interesting analysis of the degree to which en route charges condition airlines’ choices of flight routes, compared with the influence of all other direct costs (such as fuel, crew and maintenance costs). The study provides both an experimental approach and a theoretical approach, and shows that there can sometimes be convenience in avoiding certain ‘expensive’ countries, with an analysis conducted on a sample of real data from 30 ECAC members from August 2002 (flights between almost 5000 origin/destination pairs during 5 days). In the study it is pointed out that, whilst en route charges are similar in magnitude to fuel costs, they are only around half the size of maintenance, crew, and fuel costs combined. Furthermore, when delays or en-route congestion occur, the impact of route charges becomes even weaker. Another aspect they reveal is the habit of airlines to always choose the same route between a given origin/destination pair, often only because they have always acted like this. Non-rational behavior such as this is difficult to take into account

Table I
BREAKDOWN OF THE TYPICAL ROUTE-DEPENDENT COSTS OF A FLIGHT

	AEA 2003	AEA 2007	Ryanair 2008	Ryanair 2009
Fuel and Oil	35 %	50 %	57 %	65 %
Maintenance	28 %	23 %	4 %	3 %
Staff	22 %	17 %	20 %	17 %
En route charges	15 %	10 %	19 %	15 %
Total	100 %	100 %	100 %	100 %

in a mathematical model. Finally in these kinds of studies the data availability is often a problem, as airlines are often unwilling to make their cost values available (for instance airline companies generally have specific contracts with fuel suppliers, and as these can differ greatly between airlines, it is difficult to consider a significant average value). However, with some analysis and calculations, the study reveals that en route charges may play a significant role in defining the routes flown by an aircraft when a given origin and destination must be connected, and the way of charging flights for Air Navigation Services could manage the demand in the European airspace. The complexity in estimating the exact impact that en route charges have on overall flight costs, and then on route choice, should not be neglected, even though the way a flight is charged is relatively easy to compute.

To have a more precise idea of the impact of en route charges on the cost of a flight with some numerical values, we can consider the typical breakdown of flight costs. We only consider costs that change with route choice, so these are en route charge costs, fuel costs, staff costs and maintenance of the aircraft. Airport charges, depreciation, marketing and other costs are not useful for our analysis as they are independent of route choice. In Table I we report some data taken from the Annual Report of Ryanair (one of the so-called ‘low cost’ airlines, which in total have a share of around 25% of the European market [18]) and other data from the Summary Report of the Association of European Airlines (AEA) which counts national airlines and others, but excludes low cost airlines (and represents around the 50% of the European market [18])³. By setting the sum of en route charge costs, fuel costs, staff costs and maintenance costs equal to 100, we can calculate the percentage contributed by each factor. One can clearly see that en route charge costs have a significant impact. They range from 10% to 19% of the route-dependent costs of a flight, with their impact being lessened when the cost of oil is high (as was the case in 2007 for instance).

IV. BEST REACTION OF AN ANSP

We apply the Network Pricing Problem (NPP) to the case of a single ANSP, which wants to determine the charges to impose on its arcs for the next year, and which knows other ANSPs’ charges. This means finding the best reaction of the ANSP to the behavior of the system (ie. the actions of other countries’ ANSPs and network users), in order to maximize its

³These data can be found on their web sites, in [19], [20] and [18]

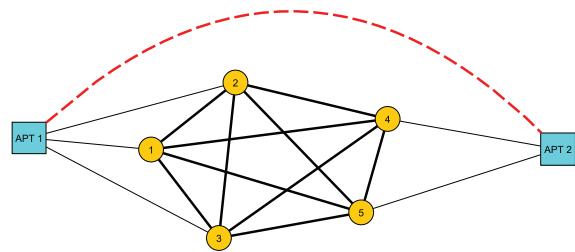


Figure 2. Example of the network

revenue. We therefore consider just one leader who wants to determine the best charge to impose on his toll arcs. All other arc costs are known. The followers are the flights which move on the network by choosing the minimum cost path. Every flight is a different commodity, and as we saw previously the relevant characteristics are not only the origin and destination, but also the operational costs, which are different for each type of aircraft and airline.

According to [1], we describe the air network for en route charges with nodes at airports and at crossing points between countries. Some considerations about the structure of this network have to be made. First of all, a country is not, in general, a convex set, as national borders tend to be highly irregular, so it may occur that a flight enters a country, exits and then re-enters it. This non-convexity property means that it may not be easy to define *a priori* an upper bound on the number of toll arcs to be used by a given flight. However it is always true that a flight does not pass through consecutive toll arcs (arcs owned by the same country). Another property of the graph which is not always valid is the completeness. The air space is divided into ‘airways’ and there may not exist an airway between every pair of nodes, both for nodes at national boundaries and for airports. During this first step, we relax these properties and make the assumptions that countries are convex and that the graph is complete. The convexity assumption allows us to say that for each flight only one toll arc can be chosen (this remains true if we consider internal airports and internal flights). The completeness assumption allows us to describe possible paths from an origin to a destination by considering all pairs between entry/exit points of a given country. Figure 2 reports an example of a network: all paths from APT1 to APT2 can be identified with all the pairs of nodes which delimit a country (e.g. one path is the pair (1, 4) which means the path APT1-1-4-APT2).

To maintain the existence of a toll free path for each flight we should consider only over flights, because if we choose a flight that lands or takes off from the country of the ANSP which we are considering, it would be obliged to pass one toll arc on entry or exit. The ANSP could therefore impose a very high (infinite) charge on his toll arcs and have a very high (infinite) revenue from this flight. In Figure 2 the toll free path is represented with a red dashed line.

A. Mathematical Model

We consider the set A of toll arcs, i.e., a flight is charged by the ANSP when passing through any arc of A . Let N be the set of all endpoints of the arcs in A . We denote as $(i, j) \in A$ the generic toll arc where both i and j belong to N . If K is the set of all the flights, the charge or toll to be paid by the generic flight $k \in K$ is equal to the product of the Unit Rate T fixed by the ANSP, the distance $l_{i,j}$ of the arc (i, j) and the factor w^k depending on the Maximum Take-Off Weight of the aircraft performing the flight. If o^k and d^k are the origin and destination points of flight $k \in K$, respectively, we denote as $d(o^k, i)$ the minimum cost path from origin o^k to node i for all $i, k \in N \times K$ and as $d(j, d^k)$ the minimum cost path from node j to destination d^k for all $j, k \in N \times K$. In this way we represent the portion of flight which is performed outside the airspace controlled by the ANSP. In addition we consider the possibility for each flight to reach its destination without crossing any arc in A . This toll free path should exist for each flight to guarantee an upper bound of the Unit Rate that the ANSP can impose on its arcs. We denote as r^k the cost of the minimum cost toll free path. We finally denote as c^k the unit cost of flight k which takes into account all other flight-related costs (e.g., fuel, maintenance and crew costs) besides the en route charges.

The Route Charges Pricing Problem (RCCP) can be written as:

$$\max_{T,x} \quad T * \left[\sum_k \sum_{i,j} x_{i,j}^k l_{i,j} w^k \right] \quad (6)$$

$$T \geq 0$$

$$\arg \min_{x,y} \quad \sum_k \left\{ \sum_{i,j} \left[d(o^k, i) + l_{i,j}(c^k + Tw^k) + d(j, d^k) \right] x_{i,j}^k \right\} \\ + \sum_k r^k y^k \quad (7)$$

$$\sum_{i,j} x_{i,j}^k + y^k = 1 \quad \forall k \in K \quad (8)$$

$$x_{i,j}^k \in \{0, 1\} \quad \forall i, j, k \in N \times N \times K$$

$$y^k \in \{0, 1\} \quad \forall k \in K$$

where T is the non-negative decision variable representing the Unit Rate fixed by the ANSP and holding on all toll arcs, $x_{i,j}^k$ is a set of binary variables equal to 1 if arc (i, j) is chosen by flight k and 0 otherwise, and y^k is a set of binary variables equal to 1 if the toll free path with cost r^k is chosen by flight k , 0 otherwise. The ANSP chooses the Unit Rate value T which maximizes its revenue (Equation 6), and knows the reaction of the followers: each flight considers all possible paths between its origin and destination, and chooses the minimum cost path (Equations 7 and 8).

B. Computational Procedure

As there is just one decision variable T at the leader level, the bilevel problem can be solved through a procedure similar

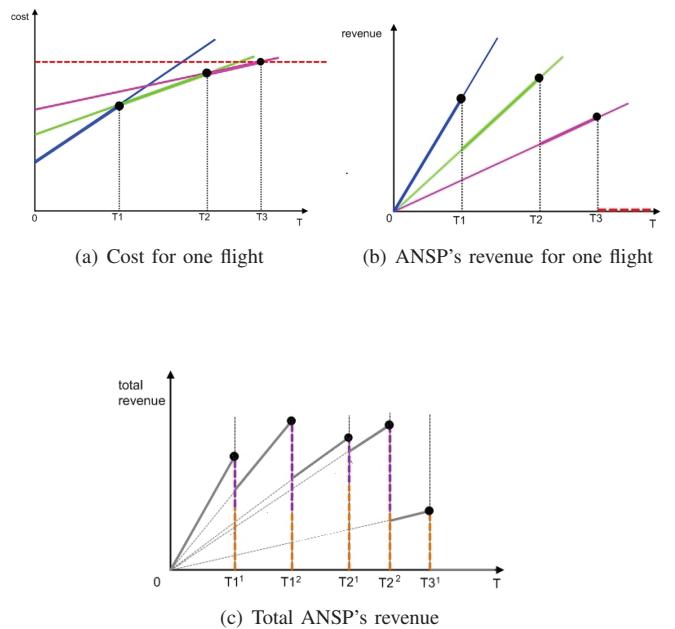


Figure 3. Computational procedure - Functions on T

to the one described for the case of a single toll arc for the NPP, which is the following one:

- 1) For each flight, we calculate the costs for all possible paths between their origin and destination (Equation 7). As the costs of the toll arcs depend on T , we identify the values of T for which the flight has convenience in changing its path choice. We obtain a piecewise linear concave function, bounded at the upper limit by the toll free path r^k , Figure 3(a).
- 2) The ANSP's revenue for a single flight is a non-continuous function, linear in each interval of T previously determined, Figure 3(b).
- 3) The above steps are repeated for each flight to find all significant T values. Finally for each T , the ANSP's total revenue is determined as the sum of the revenues from each flight, Figure 3(c). It is then straightforward to identify the Unit Rate value which maximizes the ANSP's revenues.

It is interesting to note that this procedure can be carried out even if the country is not convex. The important point for the computation described above is to know all possible paths for each flight; once we know these, even if there is more than one toll arc in any one of them, we know the total distance flown over toll arcs by the flight, and so we can proceed as described.

It may not be very easy however to find all the possible paths, as they could be great in number in a complete graph. The first stage of our algorithm requires us to enumerate all possible paths in the network for a given origin/destination pair, and in general this problem has a high degree of complexity (NP-sharp). In a real case however, due to the particular

Table II
ORIGIN/DESTINATION PAIRS OF THE EXAMPLE

Origin		Destination		Paths outside Switz.	Paths through Switz.
EDDM	Munich	LFBO	Toulouse	2	1
EGLL	London	LIPZ	Venice	1	1
EDDM	Munich	LEBL	Barcelona	2	2
EDDM	Munich	LIMC	Milan	1	3
EBBR	Brussels	LIRF	Rome	3	2
EDDF	Frankfurt	LIRF	Rome	2	2
EHAM	Amsterdam	LIRF	Rome	2	2
EIDW	Dublin	LIPZ	Venice	1	1
EIDW	Dublin	LIRF	Rome	4	2
LFPG	Paris	LOWW	Wien	3	3

topology of the air network, the problem would not be relevant as there are only a limited number of possible airway paths that a flight can use for each origin/destination pair (in general, in Europe, a flight from a given origin, can choose in between around 6-7 paths to reach its destination [2]). This non-completeness property of the real network helps to reduce the number of possible paths between two origin/destination points, and to avoid a large increase in the computational time required.

V. PRELIMINARY CASE STUDY

We delineate a preliminary case study considering a sample of flights passing through Switzerland, a central European country. In order to build the model we first need data about the network (the number and locations of entry/exit points of Swiss airspace and the length of the arcs inside and outside of Switzerland) and about the aircraft used (the aircraft weight is needed to calculate en route charges, and cost parameter values such as fuel and others are needed to calculate all route-dependent costs). We then need to know the Unit Rate values of all of the countries bordering Switzerland.

The network topology and arc distances for 10 Origin/Destination pairs have been extracted with the aid of the ‘System for traffic Assignment and Analysis at a Macroscopic level’ (SAAM) software relying on actual flight data from 29 June 2007. The pairs and the number of paths for each of pair are reported in Table II.

We choose seven different types of aircraft, which are commonly used for European flights, such as Airbus, Boeing and ATR, and then derive all their flight cost data from standard figures publicly available.

The official Unit Rates values valid in June 2007 can be easily found on the EUROCONTROL CRCO web site.

To determine the Unit Rate value which maximizes the revenues of the Swiss ANSP, we solve the RCCP with the procedure previously described. Due to lack of space, we now report in detail only the steps made for the first Origin/Destination pair (Munich-Toulouse). Figure 4 shows the map of all the paths for flights between Munich and Toulouse, and Table III reports their distance data.

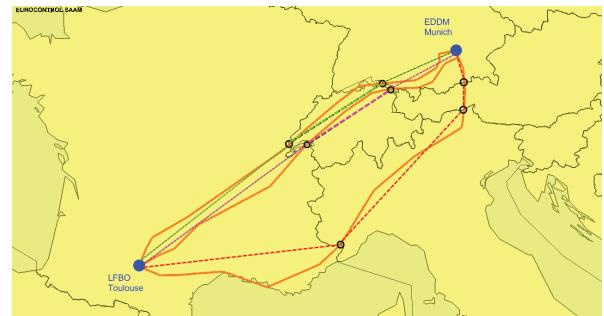


Figure 4. Routes between Munich and Toulouse

Table III
ROUTES BETWEEN MUNICH (EDDM) AND TOULOUSE (LFBO)

	DIST. FLOWN (km)	TIME FLOWN (min)	CRCO DIST. (km)
Green path	1.002,21	86,00	963,52
ED (Germany)	224,18	19,33	202,44
LS (Switzerland)	287,17	22,32	279,74
LF (France)	490,85	44,35	481,34
Pink path	996,41	83,48	970,45
ED (Germany)	218,54	24,40	204,23
LS (Switzerland)	248,19	17,33	244,41
LF (France)	529,69	41,75	521,81
Red path	1.171,26	100,76	1.110,69
ED (Germany)	84,23	8,15	82,75
LO (Austria)	68,99	5,15	68,87
LI (Italy)	469,30	36,80	451,77
LF (France)	548,75	50,65	507,30

Using these data, we are able to calculate the cost of each path as a function of the Unit Rate value T . The cost of a path is composed of a fixed part and a variable part: the fixed part is composed of fuel, maintenance and crew costs for all the distance or time flown plus the cost of en route charges for the distance flown in all countries except Switzerland; the variable part is represented by the cost of en route charges for the distance flown over Switzerland (this is the product of the Weight Factor and the Distance Factor, which are fixed for that given commodity and path, multiplied by the Unit Rate value T , which can ideally vary between zero and infinite). The paths which do not pass over Switzerland do not have a variable cost component. The product of the Weight Factor and the Distance Factor is defined as a Service Unit (SU). In Table IV we report these cost values calculated for three flights between Munich and Toulouse.

These values are reported graphically in Figures 5(a), 5(c) and 5(e) for these flights, where it is possible to see the dependence on T . From these values it is therefore possible to derive the ANSP’s revenues, which are reported in Figures 5(b), 5(d) and 5(f).

To calculate the ANSP’s revenue from these three flights, the contributions of each of them must to be summed, as reported in Figure 6: the blue function is constructed as the sum of the black functions of Figures 5(b), 5(d) and 5(f).

These calculations have been repeated for flights over all the

Table IV
COSTS OF FLIGHTS BETWEEN MUNICH AND TOULOUSE

Flight	Path	Total Fixed Costs (EUR)	Variable Cost (SU)
11: A319 flight	Green	3272,55	3,06
	Pink	3258,88	2,68
	Red	4060,67	0,00
12: B744 flight	Green	7011,04	6,69
	Pink	6991,02	5,84
	Red	8707,76	0,00
13: AT72 flight	Green	1417,10	1,72
	Pink	1413,79	1,51
	Red	1788,77	0,00

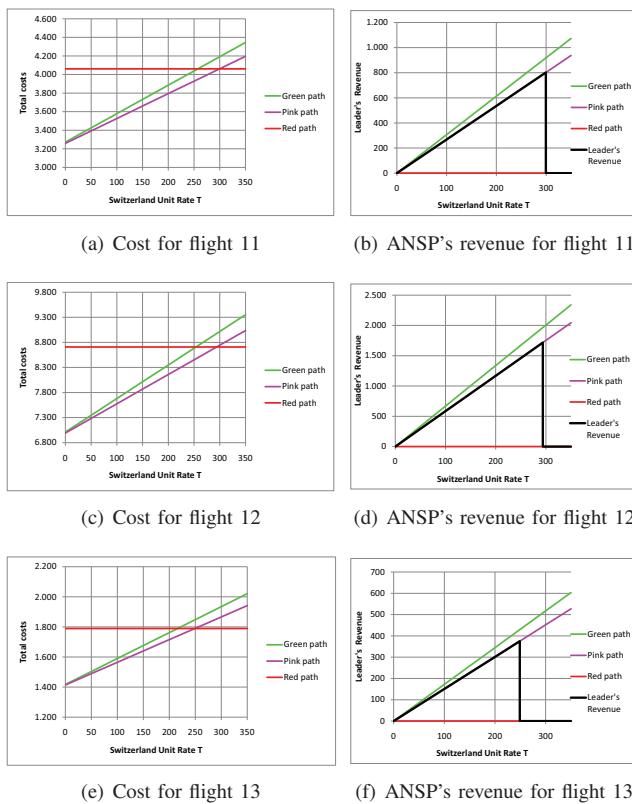
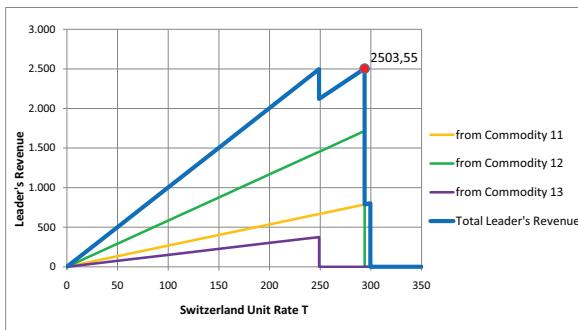
Figure 5. Flights cost and ANSP revenues - Functions on T 

Figure 6. ANSP's revenues from flights 11-12-13

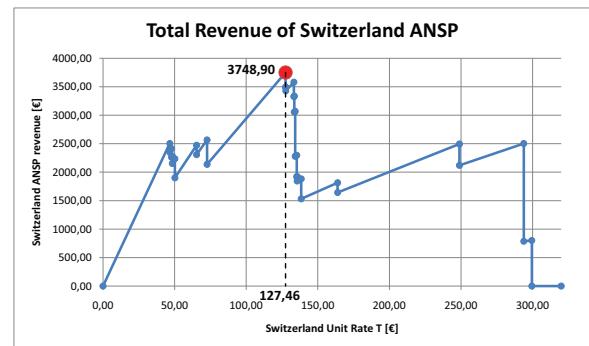


Figure 7. Preliminary example - Revenue of the ANSP

routes chosen for the example (three different types of aircraft for each pair). For certain routes it can be seen that taking a path which avoids Switzerland will always be better, for any value of T , including for $T = 0$, the minimum possible value (meaning that a path outside Switzerland is always cheaper than the fixed part of the paths inside Switzerland); in this case it clearly makes no sense for flights to pass through Switzerland. In our formulation none of these situations provide any revenue for the Swiss ANSP, and so they are not relevant. Thus in our example we have 18 flights which can choose to pass or not through Switzerland, in dependence on the Unit Rate value T fixed by the Swiss ANSP, and the ANSP's revenue has been calculated for all of them, depending on the Unit Rate value T .

Finally the ANSP's total revenue is the sum of the contributions for each flight. Based on the available flights of our preliminary example, Figure 7 shows the revenue function for the Swiss ANSP and highlights the Unit Rate value which maximizes its revenues. In this case the ANSP should fix $T = 127,46$ to achieve the maximum revenue from this sample of flights. As we made all the calculations on a small set of flights, by no means representative of the level of traffic over Europe, the result cannot be considered significant for its value, but it is interesting to gain a better understanding of the procedure and to reveal its potential on real data.

VI. CONCLUSION

This paper formulates a Network Pricing Problem addressing the case where an authority controlling a set of arcs fixes a unique value such that any commodity traversing these arcs has to pay a toll proportional to this common value. This framework depicts the way most European ANSPs are likely to behave in the near future when they determine the Unit Rate values which maximize their revenues. This is because an airline flying through an airway under the responsibility of a given ANSP has to pay it the so-called en route charges, and these charges are proportional to the Unit Rate set by the ANSP. By exploiting the structure of the problem, we propose an exact algorithm to compute the optimal Unit Rate relying on real air traffic data and realistic flight cost figures. The algorithm is polynomial except for the first step, which

enumerates all possible paths in the network for a given origin/destination pair. However, the air network has a fairly simple topology, meaning there are only a few different routes possible for each flight. Our results also suggest that the Unit Rate can indeed be an instrument for an ANSP to modify the path choice of flights. Further investigations should be carried out on a larger data set, to understand the computational time needed to solve a real scale problem.

As an unavoidable hypothesis of the NPP is to have a toll free path for each commodity, we were obliged to consider only flights over the country of study, as any flight which takes off or lands in it will not have a toll free path. However, as these flights have a significant share of the total number of flights, they are an important source of revenue for the ANSP, and it is important to consider them. Further studies should be carried out to look for a way to include flights taking off and landing in the country considered. An idea could be to fix an upper bound to the Unit Rate value, but it would not be trivial to decide the value of this upper bound.

We considered all parameters as deterministic, meaning that their precise values are known a priori. This may not always be true however: the ANSP aims to fix the Unit Rate that will be applicable to flights for the following year. It is therefore reasonable to suppose that some parameters, such as the level of traffic or the cost values, will have a degree of uncertainty around the values of the previous year. In this case a robust optimization approach could be used.

More generally, it could be interesting to consider the whole European system and the ‘competition’ between more than one ANSP, as they simultaneously fix their Unit Rates. In this case we will face a bilevel problem with multiple leaders, and thus with a game theory approach it could be possible to see if there are Nash equilibria or not, and if cooperation could bring advantages. It is straightforward to see that if we consider just one flight, or more than one but all equal (i.e. an uniform flow), the path choice is unique for each combination of tax values. As arcs are uncapacitated, the lower level problem is a shortest path problem, which is known to be linear and continuous. The equilibriums should therefore be searched in between tax values where the flight does change its best path choice. For more than one different flight the problem becomes quite complicated, as the lower level is a multicommodity minimum flow problem, which is no longer continuous and has to be solved through the integer programming. Then we need to study the variation of the flow distribution against arcs cost, and at present we are trying to find conditions guaranteeing the existence of equilibriums (or at least the non-existence).

Finally, the model we proposed for en route charges in Europe could also be generalized for other transportation problems with a similar structure, where a leader wants to fix a charge per kilometer and commodities travel on the network. We saw that this particular kind of NPP can be solved using a relatively simple procedure. It could be interesting to conduct a deeper mathematical analysis of this particular case of an NPP, to prove the computational complexity and to see if there are some useful properties (e.g. particular shapes of

objective functions or particular structures of the network that permit easy procedures, besides the small number of paths for each commodity). Moreover, a sensitivity analysis could be conducted against some parameters (for example, fixed arc costs), to quantitatively investigate the stability of the model. To our knowledge, no studies have yet been carried out to perform a sensitivity analysis on a NPP.

ACKNOWLEDGMENT

The second author acknowledges support from the Communauté Française de Belgique — Actions de Recherche Concertées (ARC).

REFERENCES

- [1] CRCO, “Customer guide to charges, version 3.5,” Eurocontrol, Tech. Rep., June 2009, <http://www.eurocontrol.int/crc0>.
- [2] L. Castelli, R. Pesenti, and W. Ukovich, “Analysis of airline operational behaviours,” in *Final Report of Work Package 4 of EUROCONTROL Care Innovative Project ‘Innovative Route Charging Schemes’*, Brussels, Belgium, 2004.
- [3] K. Button and G. McDougall, “Institutional and structure changes in air navigation service-providing organizations,” *Journal of Air Transport Management*, no. 12, pp. 236–252, 2006.
- [4] M. Labb  , P. Marcotte, and G. Savard, “A bilevel model of taxation and its application to optimal highway pricing,” *Management Science*, no. 12, pp. 1608–1622, 1998.
- [5] H. Stackelberg, *The Theory of Market Economy*. Oxford University Press, 1952.
- [6] J. Bracken and J. McGill, “Mathematical programs with optimization problems in the constraints,” *Operations Research*, no. 21, pp. 37–44, 1973.
- [7] R. Jeroslow, “The polynomial hierarchy and a simple model for competitive analysis,” *Mathematical Programming*, no. 32, pp. 146–164, 1985.
- [8] P. Hansen, B. Jaumard, and G. Savard, “A new branch-and-bound rules for linear bilevel programming,” *SIAM Journal on Scientific and Statistical Computing*, vol. 5, no. 13, pp. 1194–1217, 1992.
- [9] L. Vicente, G. Savard, and J. J  dice, “Descent approaches for quadratic bilevel programming,” *Journal of Optimization Theory and Applications*, no. 81, pp. 379–399, 1994.
- [10] B. Colson, P. Marcotte, and G. Savard, “Bilevel programming: A survey,” *4OR: A Quarterly Journal of Operations Research*, no. 3, pp. 87–105, 2005.
- [11] —, “An overview of bilevel optimization,” *Annals of Operations Research*, no. 153, pp. 235–256, 2007.
- [12] G. Heilporn, “Network pricing problems: complexity, polyhedral study and solution approaches,” Ph.D. dissertation, Universit   Libre de Bruxelles, Universit   de Montr  al, 2008.
- [13] A. F. van der Kraaij, “Pricing in networks,” Ph.D. dissertation, Proefschrift Universiteit Maastricht, 2004.
- [14] M. Raffarin, “Congestion in european airspace: A pricing solution?” *Journal of Transport Economics and Policy*, vol. 38, Part 1, pp. 109–126, January 2004.
- [15] L. Castelli, P. Debels, and W. Ukovich, “Route-charging policies for a central european cross-border upper airspace,” *Journal of Air Transport Management*, no. 11, pp. 432–441, 2005.
- [16] L. J. Basso and A. Zhang, “On the relationship between airport pricing models,” *Transportation Research Part B*, no. 42, pp. 725–735, 2008.
- [17] T. H. Oum and X. Fu, “Air transport security user charge pricing: An investigation of flat per-passenger charge vs. ad valorem user charge schemes,” *Transportation Research Part E*, no. 43, pp. 283–293, 2007.
- [18] AEA, “Summary report, operating economy of aea airlines,” Tech. Rep., 2007, <http://filesaea.be/RIG/Economics/DL/SumRep07.pdf>.
- [19] Ryanair, “Annual report,” Tech. Rep., 2008, http://www.ryanair.com/site/about/invest/docs/2008/q4_2008_doc.pdf.
- [20] —, “Annual report,” Tech. Rep., 2009, http://www.ryanair.com/site/about/invest/docs/2009/q4_2009_doc.pdf.

Track 10

Safety and Security

Bayesian Analysis of Accident Rate, Trend and Uncertainty in Commercial Aviation

Edwin A. Bloem & Henk A.P. Blom

National Aerospace Laboratory NLR

Amsterdam, The Netherlands

bloem@nlr.nl & blom@nlr.nl

Abstract— An established approach in the evaluation of aviation accident statistics is to determine point estimates of the accident rate by dividing number of accidents by number of flights and to determine an uncertainty interval through evaluation of the underlying binomial distribution. The trend, however, is not estimated. Another established approach is to perform a regression analysis to estimate rate and trend, but then uncertainty is not estimated. In this paper we overcome these limitations of established approaches by studying the problem as one of Bayesian estimation of the joint conditional density function of accident rate and trend given accident and flight statistical data. Subsequently, a particle filter is used in order to perform numerical evaluations. The novel approach is shown to work well on commercial aviation accident data.

Keywords-*Bayesian estimation; accident statistics; particle filtering; uncertainty estimation*

I. INTRODUCTION

Aviation accident data provides essential information to monitor aviation safety. If collected at large scale then this data provides insight into the progress made by aviation industry and they may indicate possible safety bottlenecks. Based on these insights, the aviation industry can set their strategy and priorities right.

A long standing problem in the evaluation of aviation accident statistics is the joint estimation of accident rate, trend and uncertainty. An established approach is to divide the number of accidents by the number of flights, and to determine a 95% uncertainty area by using the underlying binomial distribution [1]. However, this approach does not estimate trend. Another established approach in estimating rate and trend is to perform a regression analysis by which the relationship between a dependent variable and one or more independent variables is analyzed. Now the uncertainty is not estimated. Thus with established approaches, either rate and uncertainty or rate and trend are jointly estimated, but not all three.

The aim of this in this paper is to overcome the limitation of the established approaches by developing a Bayesian approach [2] towards the estimation of the joint probability density function of the accident rate and the trend. The numerical evaluation of such a Bayesian approach has become possible due to the development of powerful sequential Monte Carlo simulation techniques [3].

For the problem of estimating the joint conditional probability density function of accident rate and trend given large scale aviation accident and flight statistics, an exact Bayesian characterization is being developed first. Subsequently, a particle filter approximation is introduced in order to perform numerical evaluations. This particle filter is then used to perform joint estimation of accident rate, trend and uncertainty from commercial aviation accident data.

The paper is organized as follows. Section II formulates the mathematical problem for piecewise constant rates per year. Section III derives a recursive Bayesian characterization of the joint conditional probability density function for the rate and trend. Section IV presents a particle filter approach towards the evaluation of this joint conditional probability density function. Section V presents the results for the particle filter applied to worldwide aviation accident data. Section VI compares the new results with classical estimation results. Section VII presents results for the case where the accident data has been split into two groups, namely air related accidents and ground related accidents. Section VIII draws conclusions.

II. MATHEMATICAL FORMULATION OF THE PROBLEM

This section proposes a specific mathematical formulation of the problem addressed in this paper. In order to characterize a model for the number of accidents per flight in a year, we assume that the accident rates are piecewise constant. We also assume that observed data is available on the number of flights and on the number of accidents per year.

Let $\lambda_k \in [0,1]$ denote the accident rate per flight in year k and $a_k \in [1-\varepsilon, 1+\varepsilon]$ the accident trend in year k , $k \in [0, \dots, F]$ and assume that these two evolve according to the following model:

$$\begin{aligned} a_k &= a_{k-1} \\ \lambda_k &= a_{k-1} \lambda_{k-1} \end{aligned} \tag{1}$$

with $\lambda_0 \in [0,1]$ and $a_0 \in [1-\varepsilon, 1+\varepsilon]$. Furthermore we assume that the initial joint probability distribution p_{λ_0, a_0} is a Uniform distribution on $[0,1] \times [1-\varepsilon, 1+\varepsilon]$.

This research has been performed with support of Netherlands CAA

Let h_k denote the number of flights in year k . The accident rate Λ_k in year k is then given by

$$\Lambda_k = h_k \lambda_k \quad (2)$$

Let κ_k denote the number of accidents in year k . We assume that κ_k given Λ_k has a Poisson distribution:

$$p_{\kappa_k|\Lambda_k}(\kappa|\Lambda) = \begin{cases} \frac{\Lambda^\kappa}{\kappa!} \exp(-\Lambda) & \kappa=0,1,2,\dots \\ 0 & \text{else} \end{cases} \quad (3)$$

Bayesian estimation problem

Let $\mathcal{H}_k = \{h_1, \dots, h_k\}$ and $\mathcal{K}_k = \{\kappa_1, \dots, \kappa_k\}$ where h_k and κ_k are the observed realizations of h_k and κ_k . Given the flight statistics \mathcal{H}_F and the accident statistics \mathcal{K}_F , the problem is to characterize the joint conditional density of λ_k and a_k :

$$p_{\lambda_k, a_k | \mathcal{H}_F, \mathcal{K}_F}(\lambda, a)$$

This joint conditional density, defines estimates $\hat{\lambda}_k \triangleq E\{\lambda_k | \mathcal{H}_F, \mathcal{K}_F\}$ and $\hat{a}_k \triangleq E\{a_k | \mathcal{H}_F, \mathcal{K}_F\}$ as follows:

$$\hat{\lambda}_k = \int \int \lambda p_{\lambda_k, a_k | \mathcal{H}_F, \mathcal{K}_F}(\lambda, a) da d\lambda \quad (4a)$$

$$\hat{a}_k = \int a \left(\int p_{\lambda_k, a_k | \mathcal{H}_F, \mathcal{K}_F}(\lambda, a) d\lambda \right) da \quad (4b)$$

From the joint conditional density, 95% uncertainty intervals for λ_k can also be determined by the values $\hat{b}_{\lambda_k}^{\text{lower}}$ and $\hat{b}_{\lambda_k}^{\text{upper}}$ such that

$$\int_0^{\hat{b}_{\lambda_k}^{\text{lower}}} \left(\int_{1-\varepsilon}^{1+\varepsilon} p_{\lambda_k, a_k | \mathcal{H}_F, \mathcal{K}_F}(\lambda, a) da \right) d\lambda = 0.025$$

$$\int_{\hat{b}_{\lambda_k}^{\text{upper}}}^1 \left(\int_{1-\varepsilon}^{1+\varepsilon} p_{\lambda_k, a_k | \mathcal{H}_F, \mathcal{K}_F}(\lambda, a) da \right) d\lambda = 0.025$$

Similarly, 95% uncertainty intervals for a_k can be determined by the values $\hat{b}_{a_k}^{\text{lower}}$ and $\hat{b}_{a_k}^{\text{upper}}$ such that

$$\int_{1-\varepsilon}^{\hat{b}_{a_k}^{\text{lower}}} \left(\int_0^1 p_{\lambda_k, a_k | \mathcal{H}_F, \mathcal{K}_F}(\lambda, a) d\lambda \right) da = 0.025$$

$$\int_{\hat{b}_{a_k}^{\text{upper}}}^{1+\varepsilon} \left(\int_0^1 p_{\lambda_k, a_k | \mathcal{H}_F, \mathcal{K}_F}(\lambda, a) d\lambda \right) da = 0.025$$

III. CHARACTERIZATION OF JOINT CONDITIONAL DENSITY

In this section a recursive Bayesian characterization of the joint conditional density $p_{\lambda_k, a_k | \mathcal{H}_F, \mathcal{K}_F}(\lambda, a)$ is derived.

Applying Bayes' rule yields:

$$\begin{aligned} p_{\lambda_0, a_0 | \mathcal{K}_k, \mathcal{H}_k}(\lambda, a) &= p_{\lambda_0, a_0 | \kappa_k, \mathcal{K}_{k-1}, \mathcal{H}_k}(\lambda, a | \kappa_k) \\ &= \frac{1}{c_k} p_{\kappa_k | \lambda_0, a_0, \mathcal{K}_{k-1}, \mathcal{H}_k}(\kappa_k | \lambda, a) p_{\lambda_0, a_0 | \mathcal{K}_{k-1}, \mathcal{H}_k}(\lambda, a) \end{aligned}$$

where c_k denotes a normalising constant.

Since $p_{\lambda_0, a_0 | \mathcal{K}_{k-1}, \mathcal{H}_k}(\lambda, a)$ is independent of the number of flights in year k , thus

$$p_{\lambda_0, a_0 | \mathcal{K}_{k-1}, \mathcal{H}_{k-1}, h_k}(\lambda, a) = p_{\lambda_0, a_0 | \mathcal{K}_{k-1}, \mathcal{H}_{k-1}}(\lambda, a).$$

it follows that

$$p_{\lambda_0, a_0 | \mathcal{K}_k, \mathcal{H}_k}(\lambda, a) = \frac{1}{c_k} p_{\kappa_k | \lambda_0, a_0, \mathcal{K}_{k-1}, \mathcal{H}_k}(\kappa_k | \lambda, a) p_{\lambda_0, a_0 | \mathcal{K}_{k-1}, \mathcal{H}_{k-1}}(\lambda, a) \quad (5)$$

Using equations (1), (2) and (3) to evaluate (5) yields

$$\begin{aligned} p_{\lambda_0, a_0 | \mathcal{K}_k, \mathcal{H}_k}(\lambda, a) &= \frac{1}{c_k} \frac{(\underline{h}_k \lambda a^k)^{\underline{\kappa}_k}}{\underline{\kappa}_k !} \exp(-\underline{h}_k \lambda a^k) \cdot \\ &\quad \cdot p_{\lambda_0, a_0 | \mathcal{K}_{k-1}, \mathcal{H}_{k-1}}(\lambda, a) \end{aligned} \quad (6)$$

Repeated substitution of (6) for $k = F, k = F - 1, \dots, k = 0$, and subsequent evaluation yields

$$\begin{aligned} p_{\lambda_0, a_0 | \mathcal{K}_F, \mathcal{H}_F}(\lambda, a) &= \frac{1}{c} \left(\prod_{k=0}^F \frac{(\underline{h}_k \lambda a^k)^{\underline{\kappa}_k}}{\underline{\kappa}_k !} \exp(-\underline{h}_k \lambda a^k) \right) \cdot \\ &\quad \cdot p_{\lambda_0, a_0}(\lambda, a) \end{aligned} \quad (7)$$

where $c = c_0 \cdot c_1 \cdot \dots \cdot c_F$.

Next we characterize $p_{\lambda_k, a_k | \mathcal{K}_F, \mathcal{H}_F}(\lambda, a)$ in terms of $p_{\lambda_0, a_0 | \mathcal{K}_F, \mathcal{H}_F}(\lambda, a)$:

$$\begin{aligned} p_{\lambda_k, a_k | \mathcal{K}_F, \mathcal{H}_F}(\lambda, a) &= \int p_{\lambda_k, a_k | \lambda_0, a_0}(\lambda, a | \lambda', a') \cdot \\ &\quad \cdot p_{\lambda_0, a_0 | \mathcal{K}_F, \mathcal{H}_F}(\lambda', a') d\lambda' da' \end{aligned} \quad (8)$$

Due to (1),

$$p_{\lambda_k, a_k | \lambda_0, a_0}(\lambda, a | \lambda', a') = \delta_{[(a')^k \lambda', a']}(\lambda, a)$$

Substituting this in (8) yields:

$$p_{\lambda_k, a_k | \mathcal{H}_F, \mathcal{K}_F}(\lambda, a) = \int \delta_{[(a')^k \lambda', a']}(\lambda, a) p_{\lambda_0, a_0 | \mathcal{H}_F, \mathcal{K}_F}(\lambda', a') d\lambda' da' \quad (9)$$

IV. PARTICLE FILTERING TOWARDS ESTIMATION OF RATE, TREND AND UNCERTAINTY

In this section a particle filter approach towards the numerical evaluation of $p_{\lambda_k, a_k | \mathcal{H}_F, \mathcal{K}_F}(\lambda, a)$ is presented. This particle filter is able to provide an arbitrarily close approximation of the true Bayesian solution by increasing the number of particles. The main idea behind this particle filter is to approximate the joint conditional density of λ_0 and a_0 given \mathcal{H}_k and \mathcal{K}_k by an empirical density that is defined by a set of particles, i.e. samples from the joint conditional density with corresponding weights.

Particles are randomly drawn from the initial distribution p_{λ_0, a_0} , each with same weights. Next the particles evolve and are updated according to the underlying stochastic model and the new measurements, where for each particle its weight is adapted based on the likelihood of the measurements for that particle. For the problem at hand, the underlying stochastic model is given by equations (1), (2) and (3), the measurements are given by \mathcal{H}_k and \mathcal{K}_k , and the weights are adapted based on equation (6). With this particle filter, estimates $\hat{\lambda}_0$ and \hat{a}_0 of the model parameters λ_0 and a_0 can be obtained by simply taking the weighted average over all particles. A formal description of this particle filter reads as follows:

A *particle* is defined as a triplet $(\mu^j, \lambda_0^j, a_0^j)$, $\mu^j \in [0,1]$, $\lambda_0^j \in [0,1]$, $a_0^j \in [1-\varepsilon, 1+\varepsilon]$, $j \in [1, \dots, N]$, where N denotes the number of particles, j refers to the j^{th} particle, μ^j denotes the weight of particle j , λ_0^j denotes the expected number of flights in year 0 of particle j , and a_0^j denotes the trend parameter of particle j . With these particles the joint conditional density of λ_0 and a_0 given \mathcal{H}_k and \mathcal{K}_k can be approximated by

$$p_{\lambda_0, a_0 | \mathcal{K}_k, \mathcal{H}_k}(\lambda, a) \approx \sum_{j=1}^N \mu_k^j \delta_{[\lambda_0^j, a_0^j]}(\lambda, a)$$

The particle filter starts with an initiation step after which it cycles through the measurement update step and output equations:

Step 1: Initiation

Start with a set of N particles in $[0,1] \times [0,1] \times [1-\varepsilon, 1+\varepsilon]$ i.e.

$$\{(\mu_{k-1}^j, \lambda_0^j, a_0^j); j \in [1, N]\}$$

with $\mu_0^j = 1/N$, λ_0^j independently drawn from $p_{\lambda_0}(\lambda)$ and a_0^j independently drawn from $p_{a_0}(a)$ for each $j \in [1, \dots, N]$, e.g. both λ_0^j and a_0^j independently drawn from Uniform distributions on $[0,1]$ and $[1-\varepsilon, 1+\varepsilon]$ respectively.

Step 2: Measurement processing

Perform for $k = 1, \dots, F$ cycling through equations (10) and (11) below:

Using measurement $\underline{\mathcal{K}}_k$, determine new weights per particle,

$$\{(\mu_k^j, \lambda_0^j, a_0^j); j \in [1, N]\} \quad (10)$$

with for the new weights, using equation (6) for $k = 1, 2, \dots$:

$$\mu_k^j = \frac{1}{c_k} \frac{(h_k \lambda_0^j (a_0^j)^k)^{\underline{\mathcal{K}}_k}}{\underline{\mathcal{K}}_k !} \exp(-h_k \lambda_0^j (a_0^j)^k) \mu_{k-1}^j \quad (11)$$

with c_k a normalising constant such that $\sum_{j=1}^N \mu_k^j = 1$.

Step 3. Joint conditional density at year k :

The particle filter outputs the joint conditional density of λ_0 and a_0 given \mathcal{H}_F and \mathcal{K}_F in the form of empirical density

$$p_{\lambda_0, a_0 | \mathcal{K}_F, \mathcal{H}_F}(\lambda, a) \approx \sum_{j=1}^N \mu_F^j \delta_{[\lambda_0^j, a_0^j]}(\lambda, a) \quad (12)$$

Substitution of (12) into (9) and subsequent evaluation yields the joint conditional density of λ_k and a_k given \mathcal{H}_F and \mathcal{K}_F in the form of the empirical density

$$p_{\lambda_k, a_k | \mathcal{K}_F, \mathcal{H}_F}(\lambda, a) \approx \sum_{j=1}^N \mu_F^j \delta_{[(a_0^j)^k \lambda_0^j, a_0^j]}(\lambda, a) \quad (13)$$

Estimates $\hat{\lambda}_k$ and \hat{a}_k are obtained by calculating the weighted average over all particles:

$$\hat{\lambda}_k \approx \sum_{j=1}^N \mu_F^j \lambda_0^j (a_0^j)^k \quad (14)$$

$$\hat{a}_k \approx \sum_{j=1}^N \mu_F^j a_0^j \quad (15)$$

Step 4. In this step we determine the probability density of κ_k given \mathcal{H}_F and \mathcal{K}_F .

From the law of total probability we have:

$$p_{\kappa_k|\mathcal{K}_F, \mathcal{H}_F}(\kappa) = \int p_{\kappa_k|\lambda_k, \mathcal{K}_F, \mathcal{H}_F}(\kappa|\lambda) p_{\lambda_k|\mathcal{K}_F, \mathcal{H}_F}(\lambda) d\lambda \quad (16)$$

Together with (13) this yields

$$p_{\kappa_k|\mathcal{K}_F, \mathcal{H}_F}(\kappa) \approx \int p_{\kappa_k|\lambda_k, h_k}(\kappa|\lambda, h_k) \sum_{j=1}^N \mu_F^j \delta_{\left[\left(a_0^j\right)^k \lambda_0^j\right]}(\lambda) d\lambda \quad (17)$$

Thanks to (3) this becomes for $\kappa = 0, 1, 2, \dots$:

$$p_{\kappa_k|\mathcal{K}_F, \mathcal{H}_F}(\kappa) \approx \int \frac{(\lambda h_k)^\kappa}{\kappa!} \exp(-\lambda h_k) \sum_{j=1}^N \mu_F^j \delta_{\left[\left(a_0^j\right)^k \lambda_0^j\right]}(\lambda) d\lambda \quad (18)$$

Subsequent evaluation yields for $\kappa = 0, 1, 2, \dots$

$$p_{\kappa_k|\mathcal{K}_F, \mathcal{H}_F}(\kappa) \approx \sum_{j=1}^N \mu_F^j \frac{\left(\left(a_0^j\right)^k \lambda_0^j h_k\right)^\kappa}{\kappa!} \exp\left(-\left(a_0^j\right)^k \lambda_0^j h_k\right) \quad (19)$$

which is the equation to be used in step 4.

V. APPLICATION OF THE PARTICLE FILTER TO WORLDWIDE AVIATION ACCIDENT DATA

The particle filter of section IV is now applied to worldwide commercial aviation accident and flight data. First we explain which input data is used. Subsequently the particle filtering results are presented.

A. Input data

The aviation accident and flight data used in this paper are from [4]. The table below specifies the criteria that have been used for the selection from this database.

TABLE I. DATA SELECTION CRITERIA

Selection Criteria of accidents and flights	
Time period	1/1/1990 – 31/12/2008
Occurrence Class	Accident
Aircraft Category	Fixed Wing
Aircraft Mass group	> 5700 kg
Location of occurrence	Worldwide (not filtered)
Operation Type	Scheduled Commercial Air Transport

In figures 1 and 2 the resulting \mathcal{H}_F and \mathcal{K}_F data is visualised. Figures 1 and 2 provide the number of flights h_k

and the number of accidents κ_k for $k = 1, \dots, 19$, which corresponds with years 1990, ..., 2008. Figure 3 provides for each year the ratio between the number of accidents κ_k and the number of flights h_k .

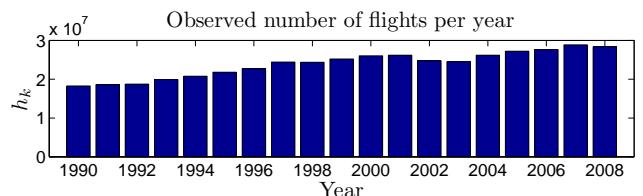


Figure 1. Observed number of flights h_k for $k = 1, \dots, 19$, corresponding with years 1990 – 2008.

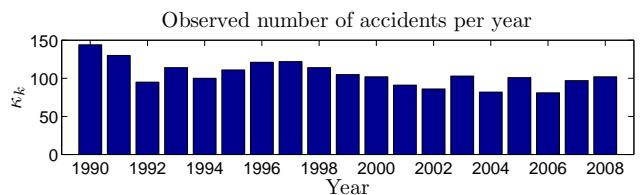


Figure 2. Observed number of accidents κ_k for $k = 1, \dots, 19$, corresponding with years 1990 – 2008.

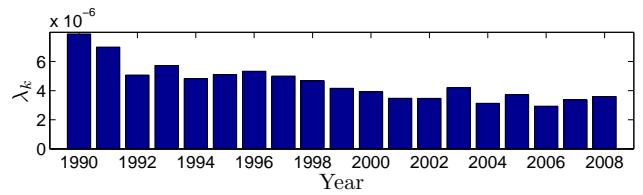


Figure 3. Ratio κ_k / h_k for $k = 1, \dots, 19$, corresponding with years 1990 – 2008.

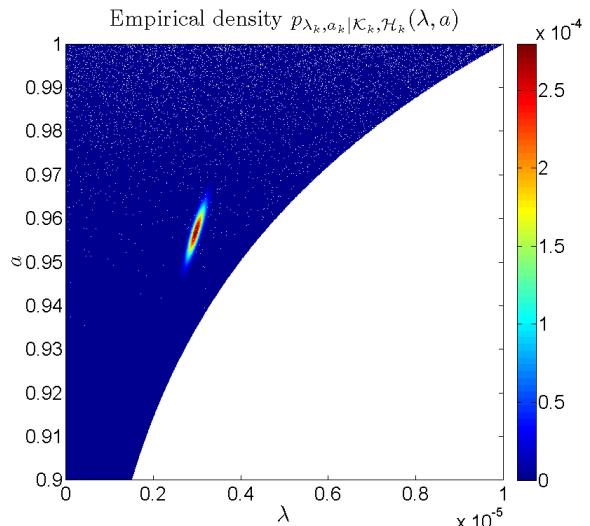


Figure 4a. Joint conditional density of $p_{\lambda_k, a_k | \mathcal{K}_k, \mathcal{H}_k}$ for $k = F$ when $\sigma = 0$.

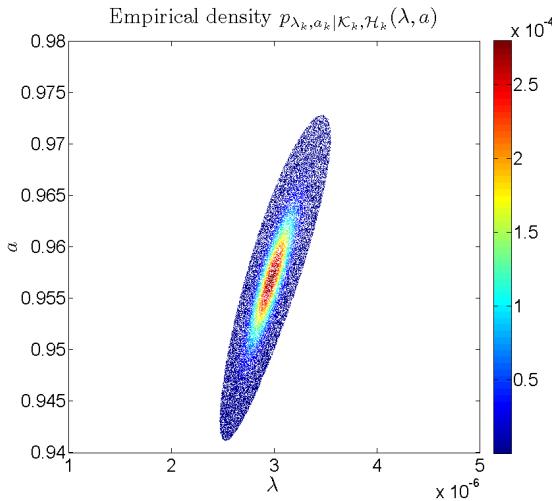


Figure 4b. Joint conditional density of $p_{\lambda_k, a_k | \mathcal{K}_k, \mathcal{H}_k}(\lambda, a)$ for $k = F$ when $\sigma=0$ and with all particles with weights below 10^{-7} ignored.

B. Particle filtering results

We use the particle filter equations of section IV with one million particles (i.e. $N = 10^6$). Particle filter based numerical evaluation of $p_{\lambda_k, a_k | \mathcal{K}_F, \mathcal{H}_F}(\lambda, a)$ is depicted in the form of the empirical joint conditional densities in figure 4. From the orientation of the empirical joint conditional densities in Figure 4 it can be observed that estimation of accident rate λ_k and accident trend a_k involves strong correlation.

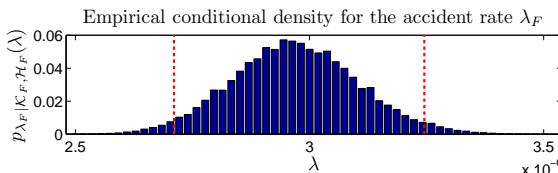


Figure 5. Empirical conditional density for the accident rate per flight in year 2008, $\lambda_F : p_{\lambda_F | \mathcal{K}_F, \mathcal{H}_F}(\lambda)$ with 95% uncertainty interval.

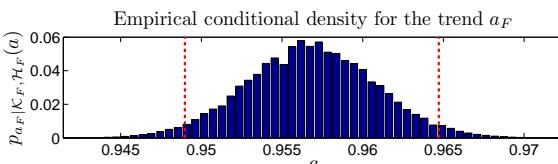


Figure 6. Empirical conditional density for the trend $a_F : p_{a_F | \mathcal{K}_F, \mathcal{H}_F}(a)$ with 95% uncertainty interval.

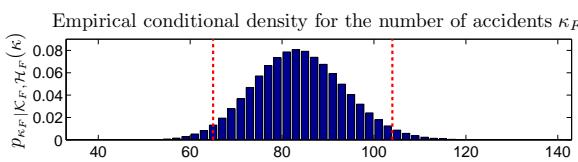


Figure 7. Empirical conditional density $p_{\kappa_F | \mathcal{K}_F, \mathcal{H}_F}(\kappa)$ for κ_F at year 2008, with 95% uncertainty interval.

Figures 5 and 6 provide marginal empirical densities for $p_{\lambda_F | \mathcal{K}_F, \mathcal{H}_F}(\lambda)$ and $p_{a_F | \mathcal{K}_F, \mathcal{H}_F}(a)$ as these resulted from the empirical joint conditional density $p_{\lambda_F, a_F | \mathcal{K}_F, \mathcal{H}_F}(\lambda, a)$.

Corresponding 95% uncertainty intervals are also given (dotted vertical lines). From the shapes of the empirical densities in these figures it can be observed that they closely resemble Gaussian densities. Figure 7 provides the marginal empirical density for $p_{\kappa_F | \mathcal{K}_F, \mathcal{H}_F}(\kappa)$ which follows from particle filter step 4. For year 2008, the estimates of the accident rate $\hat{\lambda}_F$ and accident trend \hat{a}_F are given by

$$\begin{aligned}\hat{\lambda}_F &= 2.97 \times 10^{-6} \\ \hat{a}_F &= 0.957\end{aligned}$$

The corresponding standard deviations $\hat{\sigma}_{\lambda_F}$ and $\hat{\sigma}_{a_F}$, and correlation coefficient $\hat{\rho}_{\lambda_F, a_F}$ are given by

$$\begin{aligned}\hat{\sigma}_{\lambda_F} &= 0.14 \times 10^{-6} \\ \hat{\sigma}_{a_F} &= 0.004 \\ \hat{\rho}_{\lambda_F, a_F} &= 0.87\end{aligned}$$

Hence, standard deviations amount some 5% of the estimated means, and there is strong correlation.

VI. COMPARISON WITH CLASSICAL ESTIMATION RESULTS

An established, approach, e.g. [1], is to determine for each year the ratio between the number $\underline{\kappa}_k$ of accidents and the number \underline{h}_k of flights as an indication of estimated accident rate for each year, and to use the underlying binomial distribution to determine a 95% uncertainty area around this point estimate.

Now we compare the classical estimated 95% uncertainty intervals with our new 95% uncertainty intervals that apply to $\kappa_k / \underline{h}_k$ with $p_{\kappa_k | \mathcal{K}_F, \mathcal{H}_F}(\kappa)$ as is illustrated in Figure 8 for $k = F$.

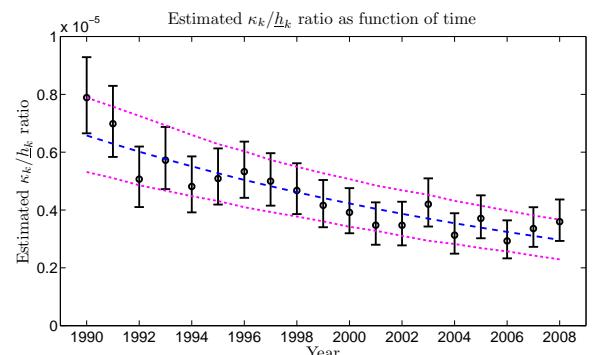


Figure 8. Newly estimated $\kappa_k / \underline{h}_k$ ratio (— = mean) with 95% uncertainty interval (----) versus classical point estimates (● ●) and 95% uncertainty interval (I).

Figure 8 shows that the sizes of the 95% areas of the newly estimated $\kappa_k / \underline{h}_k$ ratios and the classical $\kappa_k / \underline{h}_k$ ratios are remarkably similar in size. This similarity in uncertainty sizes

forms a strong indication that the novel approach yields valid results. The specific advantage of the novel approach is that it gives a much smoother estimate of the evolution of the mean over time. It should be noticed that due to (1), our novel estimates in figure 8 are based on the implicit assumption that the trend is fixed over the period considered. Similarly, the classical estimates in figure 8 are based on an implicit assumption; i.e. that the accident rates per flight in subsequent years are independent of each other.

VII. RATE, TREND AND UNCERTAINTY FOR AIR AND GROUND RELATED ACCIDENTS

We also estimate rate, trend and uncertainty separately for air related accidents and ground related accidents. The resulting estimates for the κ_k / h_k ratio are depicted in Figure 9. This shows that since 2003 the ground related accident ratio tends to overtake lead from the air related accident ratio.

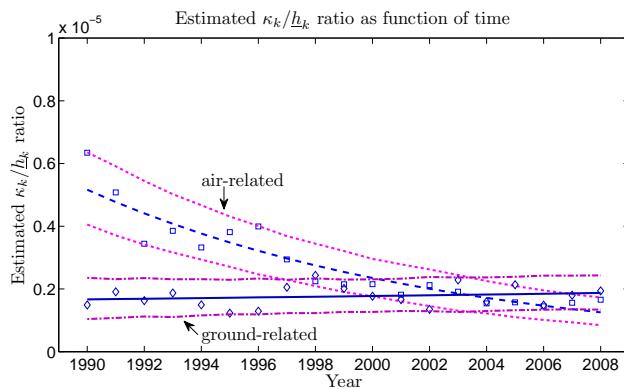


Figure 9. Newly estimated air related κ_k / h_k ratio (— — = mean) with 95% uncertainty interval (----) versus ground related κ_k / h_k ratio (— — = mean) with 95% uncertainty interval (----), with measured air related κ_k / h_k ratios (□) and ground related κ_k / h_k ratios (◊).

For year 2008, the estimates of air related accident rate $\hat{\lambda}_{F,\text{air}}$ and accident trend $\hat{a}_{F,\text{air}}$ become:

$$\begin{aligned}\hat{\lambda}_{F,\text{air}} &= 1.25 \times 10^{-6} \\ \hat{a}_{F,\text{air}} &= 0.924\end{aligned}$$

and the estimated standard deviations $\hat{\sigma}_{\lambda_F,\text{air}}$ and $\hat{\sigma}_{a_F,\text{air}}$, and correlation coefficient $\hat{\rho}_{\lambda_F,a_F,\text{air}}$ become :

$$\begin{aligned}\hat{\sigma}_{\lambda_F,\text{air}} &= 0.08 \times 10^{-6} \\ \hat{\sigma}_{a_F,\text{air}} &= 0.005 \\ \hat{\rho}_{\lambda_F,a_F,\text{air}} &= 0.90\end{aligned}$$

Hence, standard deviation for the estimated rate amount some 6% of the estimated means, and there is strong correlation. For air related accidents this means that the estimated trend value is significant. For year 2008, the estimates of ground related accident rate $\hat{\lambda}_{F,\text{ground}}$ and accident trend $\hat{a}_{F,\text{ground}}$ become:

$$\begin{aligned}\hat{\lambda}_{F,\text{ground}} &= 1.87 \times 10^{-6} \\ \hat{a}_{F,\text{ground}} &= 1.007\end{aligned}$$

and the estimated standard deviations $\hat{\sigma}_{\lambda_F,\text{ground}}$ and $\hat{\sigma}_{a_F,\text{ground}}$, and correlation coefficient $\hat{\rho}_{\lambda_F,a_F,\text{ground}}$ become

$$\begin{aligned}\hat{\sigma}_{\lambda_F,\text{ground}} &= 0.12 \times 10^{-6} \\ \hat{\sigma}_{a_F,\text{ground}} &= 0.007 \\ \hat{\rho}_{\lambda_F,a_F,\text{ground}} &= 0.84\end{aligned}$$

Again, standard deviations amount some 6% of the estimated means, and there is strong correlation. For ground related accidents this means that the estimated trend value does not deviate significantly from being at the steady value of 1.0.

The above results show that the estimated mean of ground related accident rate is in 2008 about 50% higher than the air related accident rate. Moreover, the estimated trends are: 0.65% per year increase in ground related accident rate and 7.5% per year decrease in air related accident rate. And this difference in estimated trends for air and ground related accident rates is statistically significant.

VIII. CONCLUDING REMARKS

In this paper a novel approach towards joint estimation of accident rate, trend and uncertainty in aviation accident data has been developed. This novel approach is based on the exact Bayesian estimation of the joint conditional density of the accident rate and trend given observed accident and flight data. For numerical evaluation a particle filter approximation has been developed. Numerical evaluations and comparison to classical approach shows the validity of the novel approach for joint estimation of accident rate, trend and uncertainty.

The novel method also shows that estimated air related accident rate has a decreasing trend whereas ground related accident rate has a slowly increasing trend. As a result of this, since 2003 estimated ground related accident rate tends to overtake lead from estimated airborne related accident rate. This clearly shows that there is an urgent need in developing measures that are effective in bending the trend in ground related accident rate from a yearly increase to a yearly decrease.

ACKNOWLEDGMENT

The authors would like to thank their NLR colleagues Gerard van Es and Johan Weijts for valuable discussions during the performance of this research.

REFERENCES

- [1] G. van Es. "A Review of Civil Aviation Accidents Air Traffic Management Related Accidents: 1980-1999," Proc. USA/Europe R&D Seminar on Air Traffic Management, Santa Fe, USA, 2001.
- [2] J.M. Bernardo and A.F.M. Smith, Bayesian Theory, Wiley, 1994.
- [3] A. Doucet, N. de Freitas, and N. Gordon, Sequential Monte Carlo Methods in Practice, Springer-Verlag, 2001.
- [4] NLR Air Safety Data Base, version 2009.

Stochastically and Dynamically Coloured Petri Net Model of ACAS Operations

Fedja Netjasov, Andrija Vidosavljevic, Vojin Tasic

Division of Airports and Air Traffic Safety
Faculty of Transport and Traffic Engineering
University of Belgrade, Belgrade, Serbia
{f.netjasov; a.vidosavljevic; v.tasic}@sf.bg.ac.rs

Abstract - Current international regulations and policies do not consider the effect of an airborne safety net for the analysis of safety risks. This widely accepted practice tends to create significant tension between the realization of the ambitious safety improvement targets of SESAR and NEXTGEN, and the standing regulations. In order to close this gap between SESAR and NEXTGEN requirements, and standing regulation, there is need for a systematic development of safety risk analyses of airborne safety nets within the specific ATM context, which may range from current practices to advanced ATM concepts. The aim of the research described in this paper is to make a contribution through the systematic development of an unambiguous model of TCAS II version 7, together with its interactions with pilots and ATC. The specific modelling formalism used for this is Stochastically and Dynamically Coloured Petri Nets (SDCPN). The developed SDCPN model contains the technical, human and procedural elements of ACAS operations. The SDCPN model is demonstrated to work well for a historical en-route mid-air collision event.

Keywords - ACAS, Petri Nets, Safety Risk Assessment, Safety Critical Systems

I INTRODUCTION

Airborne Collision Avoidance System (ACAS) constitutes a world-wide accepted last-resort means of reducing the risk of mid-air collision (MAC) between aircraft [1]. Key elements of the current ACAS consist of TCAS II version 7 and procedures for pilots using this system. TCAS is intended to provide last-minute collision avoidance guidance directly to the flight crew [2]. Hence, TCAS forms the last layer in the multi-layered defence against MAC, with all other layers typically belonging to ground based ATM. Although recent accidents (Überlingen, Germany, 2002; Amazon jungle, Brazil 2006) show that the current ACAS is not perfect, there are many more known examples where ACAS made a positive difference.

Current ICAO risk/safety assessment policy is restrictive relative to ACAS in the sense that maximum values for mid-air collision risk are defined under the explicit assumption that the effect of an airborne safety net is not considered. This is also the case with Eurocontrol policy, which states that safety nets in general (both airborne and ground) should not be taken into account in the risk/safety assessment process [3, 4].

Mariken Everdij, Henk Blom

Air Transport Safety Institute
National Airspace Laboratory NLR
Amsterdam, The Netherlands
{everdij, blom}@nlr.nl

In view of the SESAR and NEXTGEN objectives of increasing both capacity and safety (advances in ATM may have significant impact on the effective performance of ACAS) there simply is a need to conduct safety risk analysis of new operations, including ACAS. And this need exists, even when the inclusion of ACAS in safety regulation would not be taken up. An example is the Airborne Separation Assurance System (ASAS) as one of the new concepts whose interaction with ACAS has proven to be important from both the procedural and the human factor aspects [5, 6, 7, 8, 9]. These examples clearly show that the only way to include ACAS in the safety assessment process is through the modelling of ACAS operations.

Modelling of ACAS operations has been the subject of research since the introduction of TCAS. Many different modelling approaches with different needs have since been identified. Several approaches have emerged for verification i.e. formal analysis of complex safety-critical systems such as TCAS: Finite State Machine approach [10], State Charts [11] and Hybrid Automata [12]. In order to understand human behaviour related to TCAS, Causal analysis [13], and Timed Knowledge-based modelling and analysis [14], are applied. Finally, the necessity to examine ACAS safety is followed by development of encounter models based on Fault Tree Analysis coupled with the Monte Carlo Simulation [15], and by Markov processes coupled with Bayesian networks [16, 17]. Apart from the mentioned models, an interactive simulator InCAS was developed [18, 19] in order to replay and analyse ACAS related incidents and to learn from encounters; and a tool called Replay Interface for TCAS Alerts (RITA) was developed for ACAS training of air traffic controllers and pilots [19].

The aim of the research described in this paper is to develop a model for risk/safety assessment of ACAS operations which would allow for the assessment of the benefit of ACAS in risk reduction in current and advanced ATM. In view of this objective, the specific modelling framework used in this research is the Stochastically and Dynamically Coloured Petri Net (SDCPN) modelling formalism. The SDCPN formalism makes it possible to model a complex distributed operation in a systematic and compositional way [20], and at the same time brings powerful analysis frameworks within reach [21] and is fully embedded in the advanced safety risk assessment methodology TOPAZ [22, 23, 24].

This research has been conducted with support from the RESET project commissioned by European Commission DG TREN under the FP6 Program.

This paper is organized as follows. Section II describes the ACAS operation from the perspectives of the pilot and the air traffic controller (ATCo). Section III provides a description of TCAS II version 7. Next, Section IV explains the development of an ACAS model using the SDCPN formalism. This ACAS model covers TCAS II version 7 as well as the pilots, the controller, some other relevant equipment and the interactions between these model entities. Section V illustrates the behaviour of the new ACAS model in case of a historical MAC. Section VI draws conclusions.

II DESCRIPTION OF ACAS OPERATION

Since January 2005, ICAO mandates the use of ACAS worldwide for all aircraft with more than 19 passenger seats or with a maximum take-off weight exceeding 5,700 kg. TCAS II Version 7 is the only TCAS version that complies with ICAO Standards and Recommended Practices (SARPs) for ACAS [1, 2, 11, 25]. TCAS is designed to work autonomously, i.e. without support of the aircraft navigation equipment, and independently of the ground systems used to provide Air Traffic Control (ATC) Services [25]. Generally, TCAS monitors the airspace around the *own* aircraft and warns pilots of the presence of other aircraft, so called *intruders*, which may present a MAC *threat*. A crucial part of TCAS is a Collision Avoidance Logic, the main functions of which are [25]: tracking, traffic advisory, threat detection, resolution advisory, TCAS/TCAS coordination, advisory annunciation and performance monitoring. In order to model an ACAS operation in this research, the operation is divided into the following phases [1, 11, 25]:

A. Normal flight

In nominal situations, i.e. during normal evolution of a flight, the aircraft crew receives instructions and clearances from the Air Traffic Controller (ATCo) and is flying according to them (manually or using the autopilot). Separation assurance is the responsibility of the ATCo. TCAS is constantly surveying the surrounding airspace, by broadcasting the interrogations and receiving replays from near-by aircraft.

B. Appearance of Traffic Alert (TA)

If an aircraft comes within the range of the own aircraft, and a collision is predicted to occur within the next 20 to 48 seconds (depending on the altitude), a TA is issued, warning the flight crew by issuing the aural annunciation “Traffic, Traffic”. The mentioned aircraft is designated as “intruder”. Immediately, an icon representing the intruder aircraft on the Cockpit Display of Traffic Information (CDTI) changes its shape and colour and becomes a solid yellow circle. The crew responds to a TA by attempting to establish visual contact with the intruder aircraft as well as with other aircraft in the vicinity. The crew should not deviate from an assigned clearance given by the ATCo, and should continue to maintain or attain safe separation while reporting to the ATCo about the situation.

C. Appearance of Resolution Advisory (RA)

If the previous situation deteriorates, and a collision is predicted to occur within 15 to 35 seconds (depending on the altitude), an RA is issued. The previously mentioned “intruder”

aircraft now becomes a “threat”. The RA includes an aural annunciation in the cockpit, being “Climb, Climb” or “Descend, Descend” (depending on the situation). An icon representing the threat aircraft on the CDTI changes its shape and colour and becomes a solid red square. In addition, the icon shows the appropriate vertical rate, which should be flown in order to resolve a conflict situation. A pilot receiving an RA should disengage the autopilot and manually control the aircraft to achieve the recommended vertical rate.

If an RA occurs, the pilot flying should respond immediately by directing attention to the RA displays and manoeuvring as indicated, unless doing so will jeopardize the safe operation of the flight. By not responding to an RA, the flight crew takes responsibility for achieving safe separation. Even if an RA manoeuvre is inconsistent with the current ATC clearance, pilots are obligated to respond appropriately to the RA. Pilots are also required to report an RA occurrence, i.e. that they are responding to the RA, to the ATCo when appropriate, and to inform the ATCo of the RA deviation as soon as possible, using the defined phraseology. ATCos are advised to not issue control instructions that are contrary to the given RA. If an aircraft has begun a manoeuvre in response to an RA, the ATCo is not responsible for providing standard separation between that aircraft and other aircraft, airspace, terrain or obstructions.

D. Return to normal flight

When the RA is cleared, the flight crew get the aural annunciation “Clear of Conflict” (CoC). After that they should advise the ATCo that they are returning to their previously assigned clearance or should acknowledge any amended clearance issued, using the defined phraseology. After that, the pilot may engage the autopilot again. The ATCo resumes responsibility for standard separation if one of the following conditions is met: a) the responding aircraft has returned to its assigned altitude, the flight crew informs the ATCo that the collision avoidance manoeuvre has been completed and that standard separation has been re-established; or b) that the responding aircraft has executed an alternate clearance and that standard separation has been re-established.

III CONCEPTUAL MODEL OF TCAS II VERSION 7

As a prerequisite for developing an SDCPN model of the ACAS operation, first a conceptual model of TCAS II version 7 is developed. The development of this conceptual model is largely based on [25]. The resulting conceptual model contains models of all algorithms used for threat detection and threat resolution, and would make it possible to conduct a simulation of any encounter scenario.

A. Threat detection algorithms

In order to determine whether a collision threat exists, i.e. to issue a TA or an RA, both the range and vertical criteria must be satisfied; i.e. if one of them is not satisfied, TCAS will not issue a TA or an RA. For checking whether the range and vertical criteria are satisfied, Range tests and Altitude tests are constantly performed during an encounter. Criteria used for making a decision about TA and RA issuance depend on the Sensitivity Level (SL) (Table 1).

Table 1. Sensitivity level and threshold values [25]

Own altitude (feet)	SL	τ (seconds)		DMOD (Nm)		ALIM (feet)	
		TA	RA	TA	RA	TA	RA
<1000	2	20	N/A [*]	0.30	N/A	850	N/A
(1000-2350]	3	25	15	0.33	0.20	850	300
(2350-5000]	4	30	20	0.48	0.35	850	300
(5000-10000]	5	40	25	0.75	0.55	850	350
(10000-20000]	6	45	30	1.00	0.80	850	400
(20000-42000]	7	48	35	1.30	1.10	850	600
> 42000	7	48	35	1.30	1.10	1200	700

NA – not available

The Sensitivity Level (SL) depends of the aircraft altitude range. SL contains values for horizontal and vertical τ thresholds in case of TA or RA issuance, and dimensions for protected airspace (Distance Modification – DMOD and Altitude Limit – ALIM) which should be satisfied in case of slow closure encounters when τ threshold values are not appropriate. During an encounter, if the horizontal or vertical τ is lower than the TA threshold or if the horizontal and vertical miss distance is lower than the TA DMOD and TA ALIM respectively, then a TA is annunciated. If the situation further worsens and τ values are lower than the RA threshold or if the miss distances are lower than the RA DMOD and RA ALIM respectively, then an RA is annunciated [25].

For the purpose of range and altitude tests, aircraft are identified in a Cartesian coordinate system (Figure 1). Let x_t^i and v_t^i be the 3D position and 3D velocity of aircraft i given in expressions (1) and (2); the superscripts x and y refer to the axis system in Figure 1, and z stands for the altitude. Let θ_t^i represent an orientation velocity vector v_t^i in the horizontal plane (measured from the x axis in counter-clockwise direction, where $0 \leq \theta_t^i \leq 2\pi$) and let ψ_t^i represent the orientation of velocity vector v_t^i in the vertical plane (measured from the horizontal plane up as positive and down as negative, where $-\pi/2 \leq \psi_t^i \leq \pi/2$).

$$x_t^i = \begin{bmatrix} x_{x,t}^i \\ x_{y,t}^i \\ x_{z,t}^i \end{bmatrix} \quad (1)$$

$$v_t^i = \frac{dx_t^i}{dt} = \begin{bmatrix} v_{x,t}^i \\ v_{y,t}^i \\ v_{z,t}^i \end{bmatrix} = \begin{bmatrix} v_t^i \cos \psi_t^i \cos \theta_t^i \\ v_t^i \cos \psi_t^i \sin \theta_t^i \\ v_t^i \sin \psi_t^i \end{bmatrix} \quad (2)$$

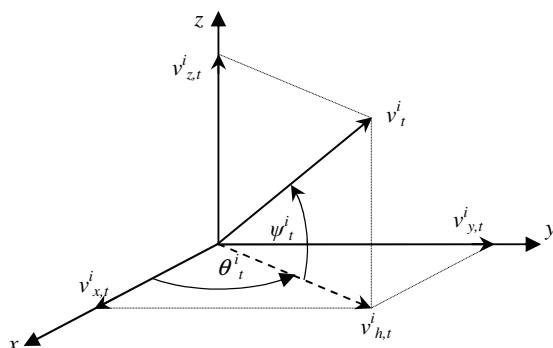


Figure 1. Velocity vector in Cartesian coordinate system

Let $x_{h,t}^{ik} = x_t^i - x_t^k$ be the distance in 3D space between own aircraft i and intruder aircraft k at time t and let $v_{h,t}^{ik} = v_t^i - v_t^k$ be the relative velocity (closing speed) between the aircraft at time t .

1) Range test:

At each moment t , both the distance and the relative velocity between own and intruder aircraft in the horizontal plane are calculated. Knowledge about both values is required in order to calculate the “time to closest point of approach” (in horizontal direction, i.e. the range τ).

Let $x_{h,t}^i = (x_{x,t}^i, x_{y,t}^i)^T$ and $v_{h,t}^i = (v_{x,t}^i, v_{y,t}^i)^T$ be the position and the velocity of aircraft i in the horizontal plane (respectively), and similarly for aircraft k . Let $x_{h,t}^{ik} = x_{h,t}^i - x_{h,t}^k$ and $v_{h,t}^{ik} = v_{h,t}^i - v_{h,t}^k$ be the horizontal distance and the relative velocity in the horizontal plane (respectively) between aircraft i and k at time t .

Define $\tau_{h,t}^{ik}$ as the time to closest point of approach (CPA) in the horizontal plane between aircraft i and k at time t , which is given by the following expression:

$$\tau_{h,t}^{ik} = \frac{-|x_{h,t}^{ik}|}{|v_{h,t}^{ik}| \cdot \cos(\delta_t^{ik} - \varphi_t^{ik})} \quad (3)$$

where δ_t^{ik} is the bearing of the velocity difference vector satisfying

$$\delta_t^{ik} = \arctan\left(\frac{v_{x,t}^{ik}}{v_{y,t}^{ik}}\right) \quad (4)$$

and φ_t^{ik} is the bearing of the position difference vector satisfying

$$\varphi_t^{ik} = \arctan\left(\frac{x_{x,t}^{ik}}{x_{y,t}^{ik}}\right) \quad (5)$$

Expression (3) is defined under the explicit condition that the denominator is not equal zero, i.e. if the following conditions are met:

$$(\delta_t^{ik} - \varphi_t^{ik} \neq \pi/2) \wedge (\delta_t^{ik} - \varphi_t^{ik} \neq -\pi/2) \wedge (|v_{h,t}^{ik}| \neq 0) \quad (6)$$

2) Altitude test:

At each moment t , both the vertical distance (separation) and the combined speed (vertical closing speed) between own and intruder aircraft are calculated. Knowledge about both values is required in order to calculate the “time to closest point of approach” (vertical τ). Let $x_{z,t}^{ik} = x_{z,t}^i - x_{z,t}^k$ and $v_{z,t}^{ik} = v_{z,t}^i - v_{z,t}^k$ be the vertical distance and the relative velocity in the vertical plane (respectively) between aircraft i and k at time t .

Define $\tau_{z,t}^{ik}$ as the time to closest point of approach (CPA) in the vertical plane between aircraft i and k at time t , which is given by the following expression:

$$\tau_{z,t}^{ik} = -\left(x_{z,t}^{ik} / v_{z,t}^{ik}\right) \quad (7)$$

Expression (7) is defined as long as $v_{z,t}^{ik} \neq 0$.

3) TA or RA issuance

The Range and Altitude tests compare given criteria (see Table 1) and calculated values for $\tau_{h,t}^{ik}$, $\tau_{z,t}^{ik}$, $x_{h,t}^{ik}$ and $x_{z,t}^{ik}$. So, whenever one of the following conditions is satisfied:

$$(0 < \tau_{h,t}^{ik} < \tau) \wedge (0 < \tau_{z,t}^{ik} < \tau) \quad (8)$$

or

$$\left(|x_{h,t}^{ik}| < DMOD \right) \wedge \left(|x_{z,t}^{ik}| < ALIM \right) \quad (9)$$

alerts shall be issued (TA or RA depending on the τ , $DMOD$ and $ALIM$ criteria given in Table 1).

B. Threat resolution algorithm

Once a threat is identified, a two-step process is followed to select the appropriate RA for the given encounter geometry. In the first step an appropriate sense is selected (upward or downward); that is, whether the aircraft needs to climb or to descend. In the second step an appropriate strength (vertical speed) is determined; that is, how rapidly the aircraft needs to change its altitude.

1) Sense selection

Let t be the moment at which an RA for own aircraft i is issued, i.e. τ_{RA} seconds remain until CPA with intruder aircraft k . The TCAS Logic makes trials with upward and downward sense for own aircraft, in order to determine which sense provides the most vertical separation at CPA (time moment $t+\tau_{RA}$ in Figure 2) under the assumption that intruder aircraft doesn't change its flight profile. The sense which provides the greatest vertical separation shall be selected.

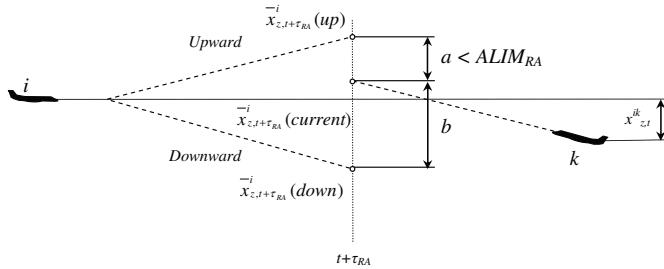


Figure 2. RA sense selection (illustrative example)

Consider a possible vertical position of aircraft i at moment $t+\tau_{RA}$ during the trial (see Figure 2):

- if the upward sense is selected

$$\bar{x}_{z,t+\tau_{RA}}^{i\downarrow}(\text{up}) = x_{z,t}^i + (v_{z,t}^i + \Delta_{z,t}^i) \cdot \tau_{RA} \quad (10)$$

- if the current rate is maintained

$$\bar{x}_{z,t+\tau_{RA}}^{i\downarrow}(\text{current}) = x_{z,t}^i + v_{z,t}^i \cdot \tau_{RA} \quad (11)$$

- if the downward sense is selected

$$\bar{x}_{z,t+\tau_{RA}}^{i\downarrow}(\text{down}) = x_{z,t}^i + (v_{z,t}^i - \Delta_{z,t}^i) \cdot \tau_{RA} \quad (12)$$

where $\Delta_{z,t}^i$ has a fixed value of 1500 feet/min [2, 11].

Two vertical separations at CPA between own aircraft i and intruder k , are recognized in the sense selection process and are given by the following expressions (see Figure 2):

$$a \equiv \left| \bar{x}_{z,t+\tau_{RA}}^{i\downarrow}(\text{up}) - \bar{x}_{z,t+\tau_{RA}}^{k\downarrow}(\text{current}) \right| \quad (13)$$

$$b \equiv \left| \bar{x}_{z,t+\tau_{RA}}^{i\downarrow}(\text{down}) - \bar{x}_{z,t+\tau_{RA}}^{k\downarrow}(\text{current}) \right| \quad (14)$$

The sense is represented by the binary variable c_t^i which takes the following values: $c_t^i = 1$ in case of the upward sense selected, $c_t^i = -1$ in case of downward sense and $c_t^i = 0$ otherwise. In case aircraft i already receives a sense from aircraft k before it has finished its own sense calculations then $c_t^i = -c_t^k$, otherwise:

$$c_t^i = \begin{cases} 1, & \begin{cases} (b > a) \vee \\ (b \leq a) \wedge \exists \varepsilon \in (0, \tau_{RA}] \text{ such that } \bar{x}_{z,t+\varepsilon}^{ik} = 0 \end{cases} \vee \\ & \begin{cases} (b > a) \wedge \exists \varepsilon \in (0, \tau_{RA}] \text{ such that } \bar{x}_{z,t+\varepsilon}^{ik} = 0 \end{cases} \wedge (a > ALIM_{RA}) \\ -1, & \begin{cases} (b \leq a) \vee \\ (b > a) \wedge \exists \varepsilon \in (0, \tau_{RA}] \text{ such that } \bar{x}_{z,t+\varepsilon}^{ik} = 0 \end{cases} \vee \\ & \begin{cases} (b \leq a) \wedge \exists \varepsilon \in (0, \tau_{RA}] \text{ such that } \bar{x}_{z,t+\varepsilon}^{ik} = 0 \end{cases} \wedge (b > ALIM_{RA}) \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

The obtained sense for the own aircraft i is coordinated through the Mode S data link with intruder aircraft k with the aim to avoid that both aircraft select the same vertical sense. So, the RA sense sent to the intruder aircraft is represented by the calculated c_t^i .

2) Strength selection

Once the sense has been selected, TCAS Logic will determine the RA strength. The RA Strength should be least disruptive to the existing flight path, while providing at least $ALIM_{RA}$ vertical separation between aircraft i and k at CPA (time moment $t+\tau_{RA}$), under the assumption that intruder aircraft doesn't change the flight profile. That means that the change of vertical speed $\Delta_{z,t}^{i*}$ should be minimal. The determination of the appropriate strength (vertical speed) should satisfy the following condition:

if $\bar{x}_{z,t+\tau_{RA}}^{i\downarrow}(\text{current}) \geq ALIM_{RA}$ then no RA is issued, otherwise the strength is calculated as follows:

$$v_{z,t}^{i*} = \begin{cases} v_{z,t}^i + \Delta_{z,t}^{i*}; & c_t^i = 1 \\ v_{z,t}^i - \Delta_{z,t}^{i*}; & c_t^i = -1 \end{cases} \quad (16)$$

where:

$$\Delta_{z,t}^{i*} = \left[ALIM_{RA} + (x_{z,t}^k + v_{z,t}^k \cdot \tau_{RA}) - (x_{z,t}^i + v_{z,t}^i \cdot \tau_{RA}) \right] / \tau_{RA} \quad (17)$$

3) Clear of Conflict annunciation

The following conditions should be satisfied in order to announce CoC and terminate the encounter: a) RAs may terminate for a number of reasons: normally, when the conflict has been resolved and the threat is diverging in range [1, 11]; b) A CoC occurs after an encounter has been resolved [11].

Let t_{CPA} be the moment when both aircraft are at CPA. Let $t' > t_{CPA}$ be the first moment when both aircraft are safely passing the CPA and the following condition is satisfied:

$$\left| \dot{x}_{h,t'}^{ik} \right| > \left| \dot{x}_{h,t_{CPA}}^{ik} \right| \quad (18)$$

then "Clear of Conflict" will be annunciated and the TCAS encounter is terminated.

IV DEVELOPMENT OF THE NEW ACAS MODEL USING SDCPN FORMALISM

In this research, ACAS operations are modelled using the Stochastically and Dynamically Coloured Petri Net (SDCPN) formalism. The main reason for using SDCPN is the possibility of modelling complex relations existing between different system elements (humans, procedures, equipment) as well as the possibility to easily determining the causes or contributing factors of non-nominal system behaviour or accidents.

Previous experiences using Petri Nets for safety analysis [26] as well as Dynamically Coloured Petri Net (DCPN) for aviation purposes [27, 28, 29] also support this choice. Once a proper ACAS model in terms of SDCPN formalism is developed, this can easily be incorporated in the modelling of an advanced ATM operation that also uses the SDCPN formalism. This way, a new ACAS model contains modules that can easily be added to some previously or future developed SDCPN modules related to current or advanced operational concepts. However, in this paper a new ACAS model is developed using SDCPN modules that work together as a standalone system.

A. Hazard identification

An important step in the TOPAZ methodology is a Hazard Identification. Once the operational concept has been sufficiently described, the hazards are identified. This is done in two steps [28]:

a) Identification of entities (agents) and their functional relationships. The agents may be humans (pilots, air traffic controllers), technical systems (e.g. navigation equipment or cockpit display, etc.), or even more abstract entities (e.g. aircraft evolution); and

b) Identification of hazards, both functional and non-functional. Hazards are best identified using dedicated brainstorm sessions with a number of participants bringing complementary expertise [28].

Because hazards could be taken from literature [2, 25] in this research no brainstorm sessions have been conducted.

B. Specification of Local Petri Nets

The SDCPN modules for the new ACAS model are developed at two hierarchical levels. The first level distinguishes the agents and the operation, where an agent is an entity that has situation awareness components. At the second level, the Local Petri Nets (LPNs) of each agent are described, where each LPN is a Petri net describing an agent-specific process. There may be connections between LPNs within the same agent or between different agents.

A Stochastically and Dynamically Coloured Petri Net is, according to [20, 22, 23, 24] given by the following tuple:

$$SDCPN = (P, T, A, N, S, C, V, W, G, D, F, I)$$

where :

P - is a set of places; T - is a set of transitions; A - is a set of arcs; N - is a node function, which maps each arc to an ordered pair of one transition and one place; S - is a set of colour types for the tokens occurring in the net; C - is a colour function, which maps each place to a colour type in S ; V and W - is a set of place-specific colour functions, which describe what happens to the colour of a token while it resides in its place; G - is a set of Boolean-valued transition guards; D - is a set of transition delays; F - is a set of (probabilistic) firing functions describing the quantity and colours of the tokens produced by the transitions at their firing; I - is an initial marking, which defines the set of tokens initially present, i.e. it specifies in which places they initially reside, and the colours they initially have.

The specification of an LPN implies determination of each element of the tuple for this LPN.

C. Agents and Local Petri Nets for ACAS operation

Each agent is represented by the multiple Local Petri Nets (LPN) mutually connected forming the SDCPN. Connections between LPNs are realised using the Compositional Specification principles presented in [20]. Five agents are recognized for the ACAS operation. They and their corresponding LPNs are given in Table 2. Interactions between agents and their corresponding LPNs are represented in Figure 3.

Table 2. Agents vs. LPN's for TCAS II Version 7 operation

Agent	LPN
Own Aircraft	
	Own aircraft state
	Own aircraft Mode S Link
	TCAS Processor
	TCAS Processor Working Mode
	CDTI Display
	CDTI Display Working Mode
	Aural Annunciation
	Aural Annunciation Working Mode
Own Aircraft Crew	
	Crew
Intruder Aircraft	
	Intruder aircraft state
	Intruder aircraft Mode S Link
Air/ground Communication Link	
	Air/ground Communication Link
Tactical Air Traffic Controller (ATCo)	
	ATCo

1) Own Aircraft as Agent

This agent contains eight LPNs and represents a technical part of the Own aircraft TCAS system. LPN *Own aircraft state_i* provides state information to LPN *Own aircraft Mode S Link_i*, which could be placed either in Work or Fail state. Through LPNs *Own aircraft state_i* and *Own aircraft Mode S Link_i*, the own aircraft and intruder aircraft positions are provided to LPN *TCAS Processor_i* which contains threat detection and threat resolution algorithms. LPN *TCAS Processor_i* can have one of the following three states (places): no conflict, conflict detection and conflict resolution. Whenever the LPN *TCAS Processor_i* is in conflict resolution state, it enables LPNs *CDTI_i* and *Aural Annunciation_i*, to move into Active state, meaning they are audio/visually representing the selected RA. LPNs *TCAS Processor Working Mode_i*, *CDTI Working Mode_i* and *Aural Annunciation Working Mode_i* represent working modes of the corresponding LPNs. TCAS Own aircraft agent is represented in Figure 4.

2) Intruder Aircraft as Agent

This agent (Figure 5) contains two LPNs and represents a technical part of the Intruder aircraft TCAS system. LPN *Intruder aircraft state_k* provides state information to LPN *Intruder aircraft Mode S Link_k* which could be placed either in Work or Fail state.

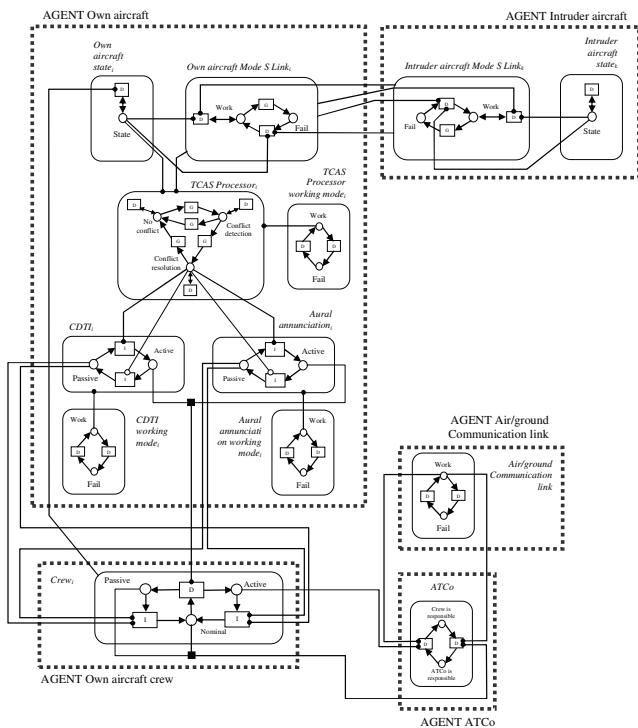


Figure 3. Interaction between agents and their corresponding LPNs for the ACAS operations

3) Own Aircraft Crew as Agent

This agent (Figure 6) contains one LPN and represents a key human entity in the ACAS operation. LPN *Crew_i* contains three places in which the crew can be: a) Nominal - in which the crew is performing their usual tasks during the flight; b) Active - in which an RA is issued and the crew is following the RA, i.e. is taking proper action in time or with some delay in

case they are too preoccupied to act immediately; and c) Passive – in which the crew is refusing to act according to the issued RA.

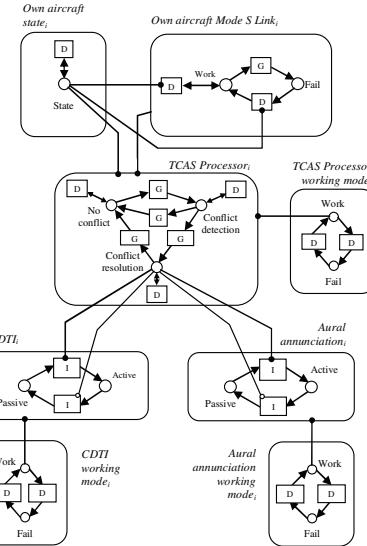


Figure 4. LPN contained in Agent Own Aircraft and their mutual relationship

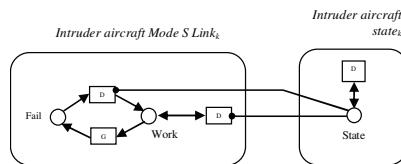


Figure 5. LPN contained in Agent Intruder Aircraft and their mutual relationship

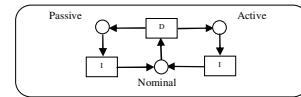


Figure 6. Agent Own Aircraft Crew

4) Air/Ground Communication Link as Agent

This agent (Figure 7) contains one LPN which represents a technical part of the system. This LPN presents working modes of the air/ground communication system.

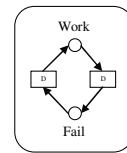


Figure 7. Agent Air/Ground Communication Link

5) Tactical Air Traffic Controller (ATCo) as Agent

This agent (Figure 8) is represented by only one LPN representing a human part of the system: LPN *ATCo* could be in one of two places: a) Crew is responsible – in which an RA is issued and the ATCo is informed about it by the Crew and the ATCo is no longer responsible for separation assurance between the aircraft in conflict; b) ATCo is responsible – in which the ATCo is responsible for separation assurance between the aircraft, or the aircraft are not in the conflict or a TA is issued.

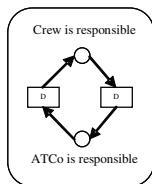


Figure 8. Agent Tactical Air Traffic Controller (ATCo)

The procedures part of the TCAS II system is represented by enabling arcs between Agent Crew and Agent ATCo (Figure 3). Therefore, whenever LPN *Crew_i* is in “Active” state, LPN ATCo switches to state “Crew is responsible”; LPN ATCo returns to state “ATCo is responsible” when an LPN *Crew_i* is in “Passive” or “Nominal” state (of course under condition that LPN Air/Ground Communication Link is in “Work” state).

V ILLUSTRATION OF THE MODEL APPLICATION

A real life accident is taken for illustration of the developed SDCPN model of ACAS operations, namely, a collision between Inex Adria DC9 and British Airways Trident 3 which occurred on September 10, 1976 over VOR Zagreb (former Yugoslavia) at FL330 [30, 31]. TCAS was not in use at the time of collision. Figure 9 provides a schematic representation of the collision location as well as flight paths of both aircraft during the few minutes before the collision [30, 31].

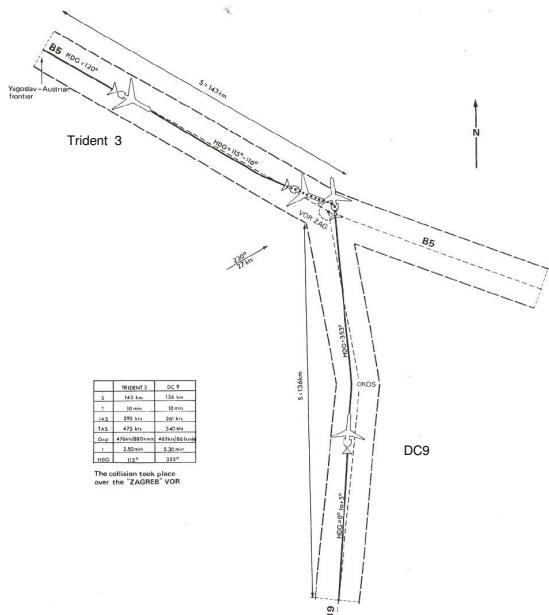


Figure 9. Schematic representation of the collision location and flight paths before collision (taken from [30])

According to the detailed vertical and horizontal situation in the last 32 seconds before collision [31] an encounter is reconstructed and input data for the simulation of the ACAS SDCPN are prepared (Table 3). Results of the ACAS SDCPN simulation are provided in Figures 10 and 11. If TCAS II would have existed at the time of the accidents, it could have prevented a collision by issuing a TA 73 sec, RA 86 sec and CoC 122 sec, from the beginning of the encounter.

Estimated minimum horizontal and vertical separations at CPA are 0.08Nm and 1933ft respectively. Own aircraft would have received a Downward sense RA while Intruder aircraft would have received an Upward sense RA.

Table 3. Encounter geometry (input)

	Own aircraft (DC9 - climbing)	Intruder aircraft (Trident 3 - cruising)
X coordinate	19.69 Nm	3.56 Nm
Y coordinate	4.54 Nm	26.78 Nm
Height	29620 ft	32960 ft
Magnetic Heading	353°	115.5°
Ground Speed	465 kt	476 kt
Vertical Speed	1670 fpm	0 fpm

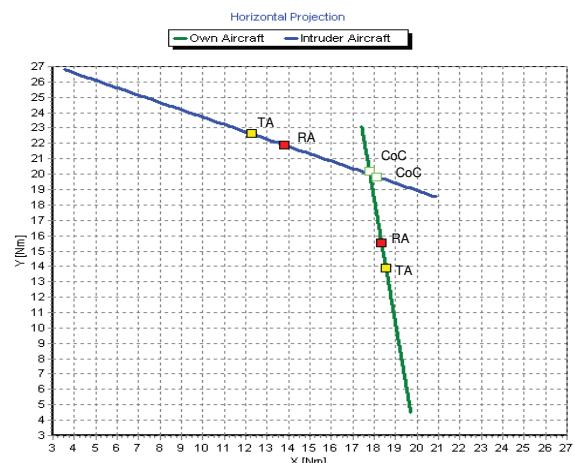


Figure 10. Horizontal situation of simulated encounter (Note: headings are not to scale)

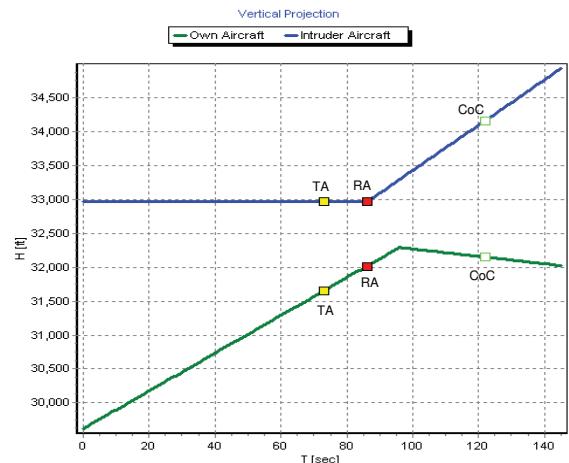


Figure 11. Vertical situation of simulated encounter (Note: rates of climb/descent are not to scale)

VI CONCLUSION

This paper presented the development of a mathematical model of ACAS operations using the SDCPN formalism. The motivation for the development of this ACAS SDCPN model is to use it in follow-up research for the safety analysis of current and advanced ATM concepts including ACAS. It was shown that the SDCPN representation is very powerful and allows the modeller to represent all elements of such a complex system (technical elements, pilots, ATCos, procedures in force), as

well as interactions between them in a flexible and modular way. An illustrative example was shown presenting the possibilities of the developed model. A further step before application in risk/safety assessment is validation of the developed SDCPN based ACAS model.

REFERENCES

- [1] International Civil Aviation Organization, "Annex 10 – Volume 4", Canada, 2002.
- [2] J. Kuchar, A. Drumm, "The Traffic Alert and Collision Avoidance System", Lincoln Laboratory Journal, Vol. 16, No. 2, 2007, pp. 277-296.
- [3] P. Brooker, "Why the Eurocontrol Safety Regulation Commission Policy on Safety Nets and Risk Assessment is Wrong", The Journal of Navigation, Vol. 57, No. 2, 2004, pp. 231-243.
- [4] P. Brooker, "Airborne Collision Avoidance Systems and Air Traffic Management Safety", The Journal of Navigation, Vol. 58, No. 1, 2005, pp. 1-16.
- [5] A. Abeloos, M. Mulder, R. van Paassen, E. Hoffman, "Potential co-operation between the TCAS and the ASAS", Proceedings of International Conference on Human-Computer Interaction in Aeronautics, France, 2000.
- [6] D. Ivanescu, D. Powell, C. Shaw, E. Hoffman, K. Zeghal, "Effect of aircraft self-merging in sequence on an airborne collision avoidance system", Proceedings of AIAA Guidance, Navigation and Control Conference and Exhibit, USA, 2004.
- [7] I. de Oliveira, P. Cugnasca, H. Blom, B. Bakker, "Modelling and Estimation of Separation Criteria for Airborne Time-Based Spacing Operation", Proceedings of 7th USA/Europe Air Traffic Management R&D Seminar, Spain, 2007.
- [8] H. Blom, B. Klein Obbink, B. Bakker, "Safety Risk Simulation of an airborne self separation concept of operation", Proceedings of 7th AIAA-ATIO Conference, Northern Ireland, 2007.
- [9] H. Blom, B. Klein Obbink, B. Bakker, "Simulated collision risk of an uncoordinated airborne self separation concept of operation", Proceedings of 7th Eurocontrol Innovative Research Workshop, France, 2008.
- [10] N. Leveson, M. Heimdahl, H. Hildreth, J. Reese, "Requirements Specification for Process-Control Systems", IEEE Transactions on Software Engineering, Vol. 20, No. 9, 1994, pp. 684-707.
- [11] Radio Technical Commission for Aeronautics, "Minimum Operational Performance Standards for Traffic Alert and Collision Avoidance System II (TCAS II) Airborne Equipment – Volume I" (RTCA/DO-185A), USA, 1997.
- [12] C. Livadas, J. Lygeros, N. Lynch, "High-Level modeling and analysis of TCAS", Proceedings of 20th IEEE Real-Time Systems Symposium, USA, 1999.
- [13] P. Ladkin, "Causal Analysis of the ACAS/TCAS Sociotechnical System", Proceedings of the 9th Australian Workshop on Safety Related Programmable Systems (SCS'04), Australia, 2004.
- [14] J. Küster-Filipe, M. Felici, S. Anderson, "Timed Knowledge-based Modelling and Analysis: On the Dependability of Socio-technical Systems", Proceedings of the 8th International Conference on Human Aspects of Advanced Manufacturing: Agility and Hybrid Automation, Italy, 2003.
- [15] J. Kuchar, "Safety Analysis Methodology for Unmanned Aerial Vehicle (UAV) Collision Avoidance Systems", Proceedings of 6th USA/Europe Air Traffic Management Research and Development Seminar, USA, 2005.
- [16] M. Kochenderfer, L. Espindle, J. Kuchar, J. Griffith, "A Comprehensive Aircraft Encounter Model of the National Airspace System", Lincoln Laboratory Journal, Volume 17, Number 2, 2008, pp. 41-53.
- [17] M. Kochenderfer, L. Espindle, M. Edwards, J. Kuchar, J. Griffith, "Airspace Encounter Models for Conventional and Unconventional Aircraft", Proceedings of 8th USA/Europe Air Traffic Management Research and Development Seminar, USA, 2009.
- [18] EUROCONTROL Experimental Centre, "InCAS 2.6 – User Guide", France, 2005.
- [19] Global Aviation Information Network (GAIN), "Guide to Methods and Tools for Safety Analysis in Air Traffic Management", USA, 2003.
- [20] M. Everdij, M. Klompstra, H. Blom, B. Klein Obbink, "Compositional Specification of a Multi-Agent System by Stochastically and Dynamically Coloured Petri Nets", in Lecture Notes in Control and Information Sciences, 337: "Stochastic Hybrid Systems: Theory and Safety Critical Application" (editors: H. Blom, J. Lygeros), Springer, 2006, pp. 325 – 350.
- [21] M. Everdij, H. Blom, "Enhancing Hybrid State Petri Nets with the Analysis Power of Stochastic Hybrid Processes", Proceedings of the 9th International Workshop on Discrete Event Systems, Sweden, 2008, pp. 400-405.
- [22] M. Everdij, H. Blom, "Petri Nets and Hybrid State Markov Processes in a Power-Hierarchy of Dependability Models", Proceedings of IFAC Conference on Analysis and Design of Hybrid System, France, 2003, pp. 355-360.
- [23] M. Everdij, H. Blom, "Modelling Hybrid State Markov Processes Through Dynamically and Stochastically And Dynamically Coloured Petri Nets", Hybridge Project, deliverable D2.4, 2005, (http://www2.nlr.nl/public/hostedsites/hybridge/documents/D2.4_Hybrid_ge-version07.pdf)
- [24] M. Everdij, H. Blom, "Piecewise deterministic Markov processes represented by dynamically coloured Petri nets", Stochastics: An International Journal of Probability and Stochastic Processes, Vol. 77, No. 1, 2005, pp. 1-29.
- [25] Department of Transportation, Federal Aviation Administration, "Introduction to TCAS II Version 7", USA, 2000.
- [26] N. Leveson, J. Stolzy, "Safety Analysis Using Petri Nets", IEEE Transactions on Software Engineering, Vol. SE-13, No. 3, 1987, pp 386-397.
- [27] J. Shortle, Y. Xie, C. Chen, G. Donohue, "Simulating Collision Probabilities of Landing Airplanes at Non-towered Airports", Simulation, Vol. 80, Issue 1, 2004, pp. 21-31.
- [28] M. Everdij, H. Blom, B. Bakker, "Modelling Lateral Spacing and Separation for Airborne Separation Assurance Using Petri Nets", Simulation, Vol. 83; Issue 5, 2007, pp. 401-414.
- [29] F. Netjasov, M. Janic, "A Review of Research on Risk and Safety Modelling in Civil Aviation", Journal of Air Transport Management, Vol. 14, Issue 4, 2008, pp. 213-220.
- [30] Department of Trade, Accident Investigation Branch, "Aircraft accident report 5/77, Report on the collision in the Zagreb Area, Yugoslavia, on 10 September 1976", UK, 1977.
- [31] Department of Trade, Accident Investigation Branch, "Aircraft accident report 9/82, Report on the collision in the Zagreb Area, Yugoslavia, on 10 September 1976", UK, 1982.

AUTHORS BIOGRAPHY

Fedja Netjasov (BS'99-MS'03) is a Teaching and Research Assistant at the Division of Airports and Air Traffic Safety, Faculty of Traffic and Transport Engineering, University of Belgrade where he received all degrees in the field of Air Transportation. He is currently finalizing his PhD thesis in the field of air traffic safety.

Andrija Vidosavljevic (BS'07) is a PhD student at the Division of Airports and Air Traffic Safety, Faculty of Traffic and Transport Engineering, University of Belgrade where he received BS degree in the field of Air Transportation.

Vojin Tasic (BS'69-MS'72-PhD'75) is a Professor and Head of the Division of Airports and Air Traffic Safety, Faculty of Traffic and Transport Engineering, University of Belgrade. He received BS degree from the same University, and MS and PhD from the University of California at Berkeley. He received BS and PhD degree in the field of Air Transportation. He is a member of the TRB Airfield and Airspace Capacity Committee and SESARJU Scientific Committee.

Mariken Everdij (MS'92-MS'94) is Senior Scientist at the Air Transport Safety Institute, National Aerospace Laboratory NLR, Amsterdam, the Netherlands. She received the first MS degree in Applied Mathematics from Twente University and the second MS degree in Industrial Mathematics from Eindhoven University of Technology. She is currently finalizing her PhD thesis on SDCPN and their associated stochastic processes.

Henk Blom (MS'78-PhD'90) is Principal Scientist at the National Aerospace Laboratory NLR, Air Transport Safety Institute, Amsterdam, the Netherlands. He received the MS degree in Electrical Engineering from Twente University, and PhD from Delft University of Technology in the field of Stochastic Control. He is an IEEE Fellow.

A Quantitative Safety Assessment Tool Based on Aircraft Actual Navigation Performance

Markus Vogel, Christoph Thiel, & Hartmut Fricke

Chair of Air Transport Technology and Logistics

Technische Universität Dresden (TUD)

Dresden, Germany

vogel@ifl.tu-dresden.de; thiel@ifl.tu-dresden.de, fricke@ifl.tu-dresden.de

Abstract: With this paper, ATM safety research at TU Dresden presents its ongoing effort to establish a safety metric for normative air traffic operations as the first step in a DFG-funded research project aimed at leveraging operational safety methodology for TMA sectors, the terminal area around major airports. After a brief introduction to the field of research and agreed-upon target levels of safety (TLS), available methodology is analyzed with the conclusion that operational safety is most appropriately assessed by probabilistically evaluating quantitative flight performance data. To gain an overall objective view on the ATM's total system safety, collision probability, a measure for the safety criticality of aircraft interactions, was selected. Though thankfully mostly marginal, collision probability is always non-zero as it arises from omnipresent navigational, flight technical, human and weather-induced positional inaccuracies. Being such an important parameter, the inaccuracies have been quantified by means of radar data analysis – leading to precise mathematical dependencies for selected critical flight phases (see Thiel & Fricke in this volume). These actual navigation performance values (ANP) are used for adaptive parameterization. As a proof of concept, segregated operations on parallel runways (SOIR) and the related planning rule (1/5 runway staggering) are subjected to safety assessment at various levels of elaboration (from a simple critical distance analysis to simulative applications of the safety assessment tool presented here). The results suggest that a collision probability approach could once have led to the planning rule, but also point out the benefits of ANP-based considerations. The paper closes with an outlook: in conjunction with an agent based simulation the tool shall help to gain insight into the safety impact of present and changed ATM procedures and variations and limitations of human performance.

Keywords: *operational air traffic safety; quantitative safety assessment; safety methodology; collision risk; human performance*

I. INTRODUCTION

The ICAO Safety Management Manual [1] defines safety as the “state [at] which the possibility of harm to persons or of property damage is reduced to, and maintained at or below, an acceptable level.” This definition entails the quantification of an acceptable level of safety (ALoS). As safety itself, the absence of unsafe conditions, cannot be measured directly, appropriate safety indicators must be defined. The yearly incident (or fatal accident) rate has always been the prime safety indicator in civil aviation because of its public visibility; its reduction or retention is the keystone safety objective. To meet this objective, regulatory and legislating bodies shall define target levels of safety (TLS) for all safety-critical systems. Various TLS for different aspects of air traffic have been defined and published in ICAO and ESARR documents (see tab. 1) specifying the highest acceptable rate of safety occurrences expressed in relative frequencies. It shall be noted that the TLS definitions use different units: they are either related to flight hours, flight segments, or typical operations.

TABLE I. LIST OF COMMONLY AGREED TARGET LEVELS OF SAFETY

Domain	Value, unit, threat	Definition
en-route operation	$5.0 \cdot 10^{-9}$ per flight hour fatal mid-air collision	ICAO ANNEX 10 [2]
surface operation	$1.0 \cdot 10^{-8}$ per operation fatal/hull-loss accident	ICAO ASMGCS [3]
approach operation	$1.0 \cdot 10^{-7}$ per approach fatal obstacle collision	ICAO PANS-OPS [4]
air traffic management	$1.55 \cdot 10^{-8}$ per flight hour direct controller contribution to reduced safety	ESARR4 [5]

The numeric values imply that safety occurrences in aviation are expected to be excessively rare. This means that low traffic volumes and brief observation times impede safety observation for existing systems. For this reason (and for the design of new systems), predictive safety assessment methodology has been developed, vastly being based on probabilistic estimation or qualitative approaches [5]. With the wide availability of quantitative data, adequate processing and description methods, and capable computing hardware, the design of quantitative safety metrics based on more complex examinations of safety indicators becomes possible and desired. Additionally, the authors see a need for a safety metric that follows the defined TLS in a way that comparable results are produced.

The research focuses the terminal area (TMA) around major airports. This sector is of special interest because traffic demand is high, driving sector load and spatial density of aircraft. Current safety regulations require separation minima to be upheld, while air traffic procedures require substantial ground guidance (controller planning and communication), rendering the TMA a bottleneck for the entire air traffic system. As the current development shows, major changes in both procedures (e.g. 4D approach trajectories) and technology (e.g. GPS navigation) drive changes in the system. These changes must be neutral or ideally beneficial to the overall system safety. With final approach and take-off being the most critical flight phases and with a proven dependency between midair collision probability and traffic density, safety in the TMA is always at stake.

Air traffic safety research at TU Dresden is ongoing since several years [11, 13] and focuses technological, operational and cognitive safety hazards. This paper presents the ongoing effort to quantitatively assess the overall system safety without inspecting individual system components. Further work shall combine the safety assessment tool with knowledge about procedures (regulations, task analyses) and human factors (cognitive effects, e.g. decision-making) to quantify the safety impact of the identified subsystems by means of an inhomogeneous, agent-based simulation.

II. SAFETY ASSESSMENT TOOL

A. Methodological Foundation

1) *Probabilistic Safety Assessment:* A substantial number of quantitative safety assessment techniques for air traffic have been developed and validated in practical applications [5]. Probabilistic approaches are predominant, building up on well established safety assessment techniques like fault and event tree analysis (FTA, ETA), probabilistic safety analysis (PSA), reliability analysis, failure mode and effects analysis (FMEA), and probabilistic cognitive models of human error. All these approaches hierarchically decompose the air traffic system and assign likelihoods of undesired behavior to the “atomic” elements of decomposition. The level of detail is commonly determined by some type of “cost” function. Total system safety is cumulated from the tree-structured safety model, resulting in an overall probability for system failure.

While practically relevant and applicable, these techniques do not address safety at normative operations but stress component failures in a binary (operational/faulty) fashion. In addition, substantial safety expertise and knowledge of the system under examination is needed for a valid assessment. As a consequence, the obtained level of safety depends on the level of detail and the amount and quality of assumptions in the underlying model.

2) *Quantitative Safety Assessment:* Air traffic safety research at TU Dresden follows a related but essentially different approach. By coupling quantitative models to a slim probabilistic safety assessment, system analysis and safety assessment are separated. There are two major benefits: (1)

openness to alternate safety metrics which can be applied to the quantitative models, (2) the ability to assess safety for existing and simulated systems alike. Both meet our research hypotheses: the authors question the sufficiency of one single safety metric for all cases and disciplines (consider technical system failure vs. human error) and deem quantitative models incorporating the findings in aviation and human factors to be superior to established safety models by making the achieved level of detail explicit (in structure and quantity). In the case of existing systems, measured flight track data is used as input for safety assessment, resulting in the highest possible level of detail.

The obvious approach for safety assessment using aircraft locations follows the well established separation concept. Time/distance- and density-based safety metrics are widespread [5, 6-8] but lack fidelity [9]. For this reason, probabilistic approaches quantifying the interaction of aircraft pairs have been investigated by numerous researchers over the past twenty years [10-12]. Overall, two classes of probabilistic functions have been addressed: location probability and conflict probes. The former addresses the discrepancy between intended and true aircraft location in dependence of various factors, e.g. steering accuracy, navigation tolerance, and weather influence. Location probability functions are of descriptive nature. The latter represent functions extending lobes into the considered aircraft’ intended trajectory. Evaluating these functions leads to safety zones weighted with criticality (time to go for evasive action such as applied in the standardized TCAS logic [2, 14]). Thus, probabilistic conflict probes are of predictive nature. The primary applications have been collision avoidance and conflict detection (e.g. TCAS, STCA).

B. Previous Work

Driven by the wide availability of radar and simulation data, previous work at TU Dresden applied location probability functions to assess safety [11, 13]. Initially, tracking inaccuracies along flight paths were modeled with tubular location probability functions aiming to support the airspace layout process with a local collision-based safety metric [11]. The concept was later extended to three-dimensional location probability functions to evaluate explicitly the collision risk at a given time and location, thereby adding temporal resolution to the spatial resolution. This resulted in a software package called *Safety Korrelator*, a Java-based tool for collision risk assessment.

Within the INTEGRA research project, Eurocontrol implemented a similar metric [14]. With the aim of developing a safety metric that facilitates comparisons between different traffic situations, *propensity*, defined as the “likelihood of a safety significant event occurring during normal operations”, was conceived. Propensity quantifies pair wise aircraft interaction in a probabilistic fashion, taking into account the normative capabilities of the air traffic control system in use. Although these capabilities are generally user-definable, [14] provides three sets of “reasonable default values” for navigational variances depending on the mode of en-route traffic control (manual, 3D, and 4D). These variances are weighted with several modifiers to reflect influences of severe weather,

traffic complexity, advanced navigation aids, and task-load induced changes in operator performance.

C. Safety Assessment Concept

Building up on previous work at TU Dresden and Eurocontrol, the core of our safety assessment concept is the evaluation of the midair collision probability that arises from aircraft location inaccuracies. The intended locations are the quantitative input into safety assessment. The system-wide collision probability is evaluated through numeric integration of all aircraft location probability density functions. Mathematically, this is less elegant (and less efficient) than the possible symbolic integration – but scientifically interesting, because spatial resolution is achieved, allowing visualization and hotspot analyses of the results.

Location inaccuracies are modeled in form of three-dimensional normal distributions, following numerous practical observations [16]. The important novelty of our approach is the integrated adaptive parameterization with actual navigation performance values (ANP) obtained from radar data analyses at major airports in Europe [17, 18 – also in this volume]. The respective mathematical analyses confirm that tracking deviations are indeed normally distributed and quantify navigation tolerances. For selected critical flight phases in the TMA, e.g. final approach and take-off, precise mathematical dependencies were derived. For other flight phases in the TMA, e.g. holding and transitions, static ANP values were determined. The quantitative results summarize all tracking deviations including flight technical, navigational and human error as well as weather influence.

The concept focuses a slim, technologically founded, and meaningful safety metric. In fact, the influence of human error on the ANP values actually contradicts this principle. Similarly, the weighting factors in INTEGRA add a substantial number of degrees of freedom to the safety metric, altering the level of detail and leaving unnecessary room for assumptions (the “use of advanced tools” weighting factor defaults to either 1.0 or 0.6 [14]). While weather influences, flight technical, and navigational errors are rightfully included in the safety assessment concept, the authors believe that the influence of human factors should be limited to the lowest possible extent in order to shift the modeling from safety assessment to the underlying quantitative model. The ANP values used for safety assessment fully exclude controller influence (while maintaining pilot influence) due to the fine-grained analysis of track data, which isolated flight segments in a way that controller intervention was out of the question. Consequently, alternate safety assessment approaches for safety assessment addressing the issue of human factors can still be incorporated. The authors rightfully acknowledge human factors influenced complexity metrics like *Dynamic Density* [8] with the note that the transfer function between controller task load and air traffic system safety is yet to be identified [9].

D. Software Implementation

With regard to code flexibility and future integration of related approaches and tools in the light of changing research objectives, the Java programming language was selected for implementation. The *safety assessment* tool itself is packaged

independently but uses the *data model* package representing the traffic scenario (airspace plus aircraft locations over time), thus serving as an intermediate information structure. The contained quantitative information is either real flight track data (e.g. from radar analyses), simulated flight track data (e.g. from real-time simulations with humans in the loop), or extrapolated track data (e.g. from computer simulations). Our software abstracts the various data sources with the *traffic generation* package, containing file readers or interfaces to externally implemented simulations. Persistence of output data is currently reached through logging.

Though runtime was considerably lowered in comparison to *Safety Korrelator*, the numeric integration of the collision probability remains computationally demanding. Various optimizations, including multi-stage processing, look-up tables for trigonometric functions, and fine-tuning of the calculation algorithm are accountable for the majority of speedup, while parallelization of the core calculation yielded only a minor speedup of roughly 1.45 (on dual-core commodity personal computers).

E. Reference Results

As a reference, levels of location and collision probability were calculated and successfully tested in their Gaussian distribution. Furthermore, the mathematical theorem, stating that the convolution of two Gaussian functions is again a Gaussian was shown to be applicable to typical aircraft interactions. As expected from commonly agreed-on TLS values, collision probabilities approach zero even at considerably unsafe aircraft separation. Therefore, logarithmic scales are used in all further diagrams showing collision probability results.

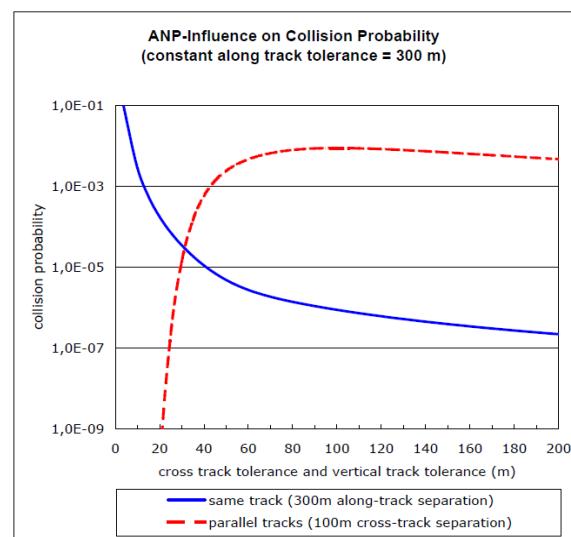


Figure 1. Influence of track keeping capability on collision probability

Next, a sensitivity analysis to the aircraft’s track keeping capability was undertaken. Although in reality, altitude measurements surpass location measurements in accuracy, both cross track (XTT) and vertical track (VTT) tolerances were jointly varied between 0 and 200 m while maintaining a constant along track tolerance (ATT) of 300 m to model the current

trend of increasing tracking performance versus constant inaccuracies in lateral positioning (e.g. self-separation). To capture severe reactions of the safety metric, aircraft were positioned in ANP-proximity (1) in line on the same track (300 m separation) and (2) in parallel (100 m separation). The results (shown in fig. 1) illustrate the nonlinearity of collision probability: Increased track keeping capability (low XTT/VTT) lowers the collision probability of laterally separated aircraft (parallel tracks, dashed red plot) but increases the collision probability of longitudinally separated aircraft (same track, solid blue plot).

III. SEGREGATED OPERATIONS ON PARALLEL INSTRUMENT RUNWAYS

A. Introduction

As a proof-of-concept, simultaneous operations on parallel or near-parallel instrument runways (SOIR, [19]), were selected for further analysis. SOIR, especially segregated operations (mode 4), is of particular interest because the father document, ICAO Annex 14 [20] gives detailed information about requirements in airport layout and procedures, yet does not reveal the underlying scientific or empirical base. The narrow airspace in focus and the high detail in the requirements permit a straightforward construction of scenarios.

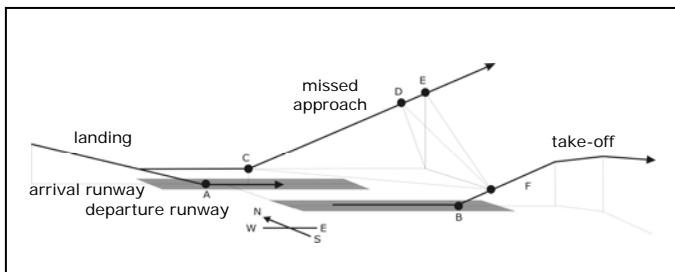


Figure 2. SOIR scenario with points of interest for minimal distances

Distinct feature of SOIR is its rule about runway spacing, allowing a reduction down to 760 m for dependent or segregated parallel operations (compared to 1035 m for independent parallel operations). For segregated operations, i.e. dedicating one runway for landings and one runway for take-offs, the spacing varies as runways thresholds are staggered relative to each other, following the so-called 1/5-rule: Separation “may be decreased by 30 m for each 150 m that the arrival runway is staggered towards the arriving aircraft, to a minimum of 300 m” ([19] pp. 4-1) and must be increased vice-versa (see fig. 2 for illustration). Three basic types of operation were extracted from [19]: (1) successful landings, (2) missed approaches initiated at the minimum descent altitude (MDA or decision height, DH) and (3) departures (see fig. 3). Touch-and-go operations were considered as well, but discarded because they do not qualify as standard operating procedures.

Asking for the foundation behind this simple-looking linear dependency, an exemplary safety analysis using our safety assessment tool is undertaken. Considering the maturity of the rule, the first hypothesis is that a separation-based approach determined the dependency. Following the idea of safety targets in aviation (TLS), the second hypothesis follows the

conjecture that a level of safety in the magnitude between enroute operations and precision approach (10^{-9} - 10^{-7} per operation) should be observable at segregated runways.

To get some methodological insight, the analyses are incrementally elaborated. First, critical distances representing selected separation minima are analyzed manually (sect. B). Then, resulting collision probabilities for these isolated cases are calculated using our algorithm (sect. C). In the next step, average values per landing operation are calculated by means of a simple simulation combined with the safety assessment tool to complement the results (sect. D). Finally, the simulation is extended with stochastic parameter variation towards a Monte-Carlo style simulation (sect. E).

B. Minimum Separation (Identification of Critical Cases)

Although obstacle collisions are without doubt the major hazard during final approach, the following analyses are constricted to airborne vehicles. Nevertheless, the case of stationary threats is covered by our approach as well: obstacles can easily be modeled as stationary conflict partners with no positional inaccuracy.

Following the separation hypothesis, critical distances are identified through geometric analysis. Certain parameters (touch-down and take-off locations, the decision height, go-around and take-off gradients) are assumed to be fixed. In particular, the nominal touch-down and take-off zones (TDZ, TOZ, [4]) are reduced to the fixed points A and B (see fig. 3). The north runway is designated for arrivals from the west, approaching with a 3° glide slope. If the approach is aborted at the DH of 200 ft, the missed approach path, an 800m horizontal section followed by a 10° incline, is taken [4]. The south runway is designated for departing aircraft (10° climb angle). The departure track diverts 30° from the missed-approach path after a brief straight-out climb phase according to [4, 19]. Without considering the following aircraft (let arriving and departing aircraft be sufficiently separated within), the critical distance for landings is easily identified as the distance between the points A and B. The critical case for missed approaches is not as easily identified. Due to identical climb angles, the missed approach path parallels the straight-out departure section. Thus, the critical distance equals the slant distance between the flight segments (D-F in fig. 3). The cases of horizontal (C-F) and vertical (E-F) line-ups are equally considered with reference to the nonlinearities described in chapter II.E.

The resulting separation minima are shown in fig. 3. Variations in runway placing (offset vs. spacing) are undertaken for the arrival runway following the 1/5 rule (thin solid black plot) while the departure runway threshold remains fixed at the origin. The two x-axes reflect this dependency (as the y-axis is used for collision probability in the succeeding charts). Vertical separator lines indicate notable points for the 1/5 rule: below –2300 m offset, runway spacing is constant at the defined minimum of 300 m. Then, the linear dependency applies. At 0 m offset, the runway thresholds are aligned. Forbidden runway placing consequently leads to unsafe separation according to the rule; this region is shaded in flat red and denoted “unsafe per definition”.

Let us discuss the separation minimum of landing vs. departing flights first (A-B in fig. 2, dotted blue plot). Point A moves along with the arrival runway as the staggering is varied according to the 1/5 rule whereas point B remains fixed. The resulting change in distance between the two points leads to a hyperbolic dependency with a clear minimum. In contrast, the separation minimum of flights aborting their approach and departing flights (D-F in fig. 2, dashed red plot) shows a different characteristic. Due to the parallel inclining path segments the influences of runway offset and runway spacing counteract, spreading the hyperbolic curve. The contribution of both runway spacing to the separation minimum also shows in the bend occurring at -2300 m offset where runway spacing reaches the defined minimum of 300 m. The other distances of interest (C-F and E-F) are omitted because they do not add any information (the difference is negligible).

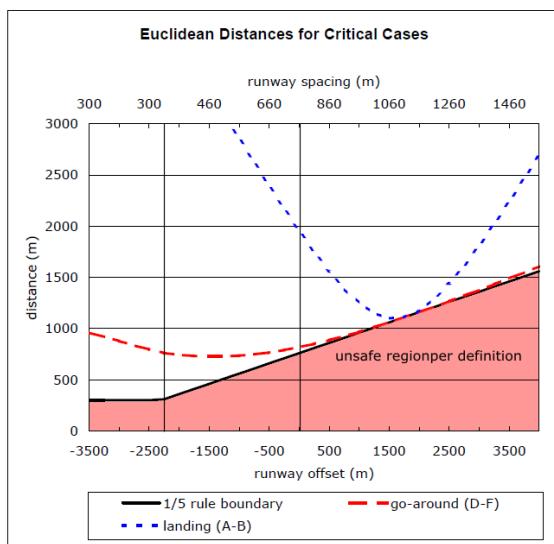


Figure 3. Separation minima for critical cases
(in dependence of runway staggering)

We conclude that distance-based considerations do not reveal a separation-based reasoning behind the 1/5 rule. The forbidden region for runway placing as well as the resulting separation minima for typical operations on SOIR both show a strong dependency on runway staggering.

C. Maximum Collision Probabilities

Subsequently, collision probabilities are calculated according to the identified separation minima in dependence of runway staggering (compare fig. 3). Utilizing our safety assessment tool, aircraft were instantiated at the points A&B (for landing operations) and D&F respectively (for go-around operations). The calculation is parameterized with arbitrary ANP values (0.2 NM XTT and VTT, 0.3 NM ATT) derived from [13, 14]. As the runway staggering is varied, the points A&D and the associated aircraft move with it, effectively altering the Gaussian location probability functions. It is important to note that the aircraft themselves remain fixed at the previously identified location of minimal separation. The results are depicted in fig. 5, maintaining the x-axis, the color-coding of plots and the shading of the “unsafe region”.

Most notable, all resulting curves are flipped and smoothed. As collision probability is a measure of risk, higher values are worse (the “unsafe region” reflects this). The smoothing happens due to the nonlinear nature of the calculation and the succeeding logarithmic scaling.

It now becomes evident that go-around operations (dashed red plot) are on the verge of the “safe” zone (flat red shading) if runways are placed along the rule’s boundary (thin solid black plot, compare fig. 3). In comparison, landing operations (dotted blue plot) are generally safer due to mostly greater separation. Nevertheless, the collision risk is in the magnitude of go-around operations at the known separation minimum at 1800 m offset (where points A and B almost align, compare fig. 2 and fig. 3). The alternate setups (C-F and E-F) do not provide any further insight and are finally abandoned.

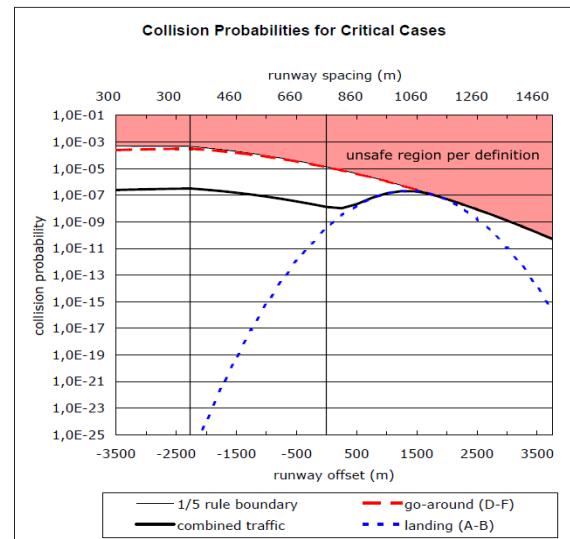


Figure 4. Collision probabilities for critical cases
(in dependence of runway staggering)

A clear indication for the foundation of the 1/5 rule can again not be concluded. A look into the frequency of operations occurring at real airports may provide further insight: missed approaches are quite rare compared to successful landings and should not influence the design of specifications as much as nominal operations. For this reason, both cases were joined with an arbitrary (but realistic) ratio of one go-around per one thousand successful landings (thick solid black plot). With this reasoning, it is possible to conclude that the 1/5 rule does have a somewhat stabilizing effect on a measure of safety (average peak collision probability per operation), though the validity of averaging peak values is doubtful.

In retrospect, the authors find it important to note that the identification of separation-critical distances is quite tedious and not trivial. Even for the simple example presented here, important aspects were initially overlooked: e.g. the fact that points A and B cease defining the minimal separation above 1800 m offset (manually corrected for combined traffic). Although a complete analysis is possible for this simplified SOIR example, more complex scenarios will be very demanding – an analytic approach will effectively lie beyond feasibility.

D. Deterministic Simulation

With a simulation of the scenario (fig. 2), peak collision probabilities can be obtained automatically. In addition, it is possible to define and calculate valid and meaningful average values (e.g. average collision probability per flight operation) that are directly comparable to the defined TLS value for an obstacle collision during approach [4].

Highly abstracting real behavior, aircraft were designed as point mass elements that follow a sequence of waypoints in a strictly linear fashion (no realistic turning behaviour). The self-contained software agents utilize a small set of flight technical parameters that were extracted from the A320 flight manual [21] as a substitute for medium-sized aircraft (according to ICAO Wake Turbulence Category) predominating traffic at major European airports. Later modeling efforts shall incorporate various aircraft sizes and proper flight behavior. The simulation itself is designed to cover all combinations of approaching and departing aircraft for each runway placing. This is achieved by deterministically delaying the arriving aircraft each time a departing aircraft enters the scene. The simulation ends when the starting configuration is reached again. The temporal resolution was chosen to be 4 s. The runway staggering was varied in steps of 250 m runway offset.

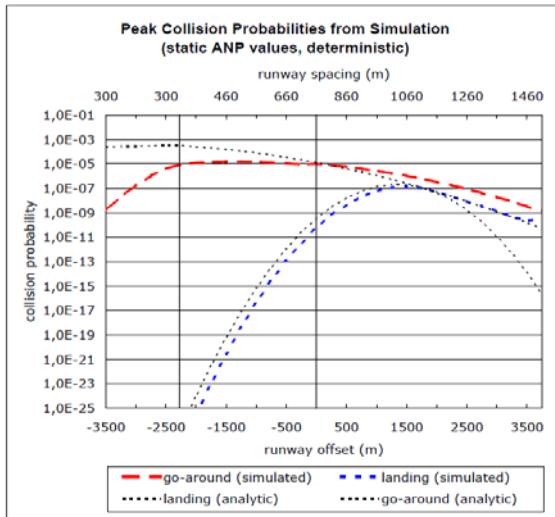


Figure 5. Maximum collision probabilities from simulation run

The peak collision probabilities from simulation are overlaid with those obtained previously in fig. 5, showing that the simulation sufficiently (re-) produces the peak values. While the plot for landing operations (dotted blue) matches the previous results (including the changed dependency after 1800 m) quite well, the plot for go-around operations (dashed red) does indicate differences. Left of zero offset, the graph stays below the expected results because the straight-out climb segment is not of infinite length (as wrongly assumed before). Right of zero offset, an undesired modeling effect raises the curve slightly: when departing aircraft turn, their probability functions extend into the missed approach path because the shape of their location probability functions does not bend with the flown trajectory, as it ideally should. (The impact on the average values is small because the effect is rare).

The average collision probabilities are shown in fig. 6. Note the change in magnitude, as the values relate to approach operations now. In comparison to the peak values, go-around operations show a changed characteristic because the exposure time to high risks comes into play now. The curve for combined operations remarkably stays within the assumed TLS range of $1 \cdot 10^{-9}$ to $1 \cdot 10^{-7}$.

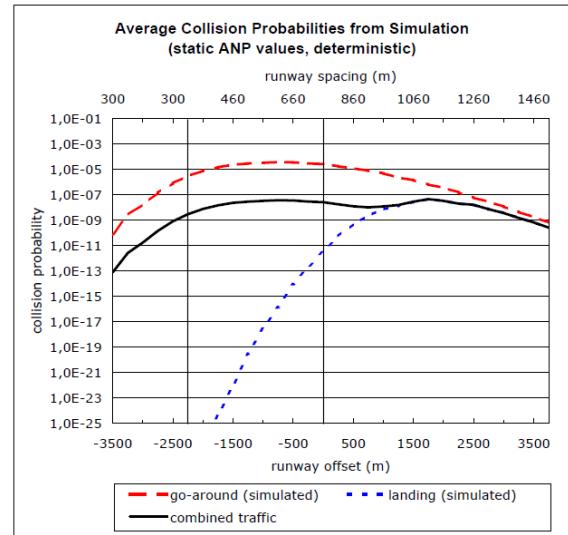


Figure 6. Average collision probabilities from simulation run

To illustrate the benefits of the numeric algorithm (namely the spatial resolution of results), fig. 7 shows a heat map obtained from the simulation for one instance of runway placing (annotations by hand; not exactly to scale).

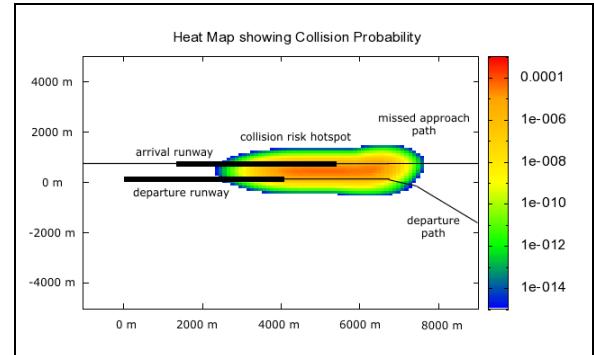


Figure 7. Heat map from simulation

E. Monte-Carlo Simulation

Monte-Carlo style stochastic parameter variation was added to recreate more vivid traffic patterns. Tab. 2 shows parameters and their variation (equal distribution). To compensate for the stochastic influence in the simulations, trial runs were performed to determine the minimal simulation time for stable results (5 hours of continuous landings or missed approaches respectively). In a first step, safety assessment of flight tracks was performed with identical ANP values to generate comparable results. In a second step, safety assessment was repeated with dynamic ANP values (compare sect. II.C).

TABLE II. SIMULATION PARAMETERS AND THEIR VARIATION

Parameter	Value	Variation
landing speed (v_{ref})	120 knots	none
take-off speed (v_2)	160 knots	none
aircraft separation	3 NM	+0..20 s
touch-down location (TDZ representative)	450 m from threshold	none
decision height (missed approach)	200 ft altitude	+0..100ft
acceleration and rotation phase (take-off)	1500 m from threshold	+0..2000 m
straight climb out phase (take-off)	1000 m	+0..1000 m

The results for constant (“static”) ANP values are shown in fig. 8. As expected, the parameter variation, which mostly spreads formerly fixed waypoint locations in the direction of runway offsets, leads to broader curves for both landings and missed approaches. While both curves maintain their shape with parameter variation, landing operations (dotted blue plot) are affected more. The reason is simple: the critical distance (A-B) is directly affected by the parameter variation in the take-off location. For missed approaches, the flight tracks between conflicting aircraft are roughly parallel, which lessens this influence. The graph for combined operations (solid black) now is somewhat constant in the familiar TLS magnitude of $1 \cdot 10^{-7}$. Though the simulative approach and the assumptions are complex and certainly not generic, the authors see some indication for the validity of the 1/5 rule in this result.

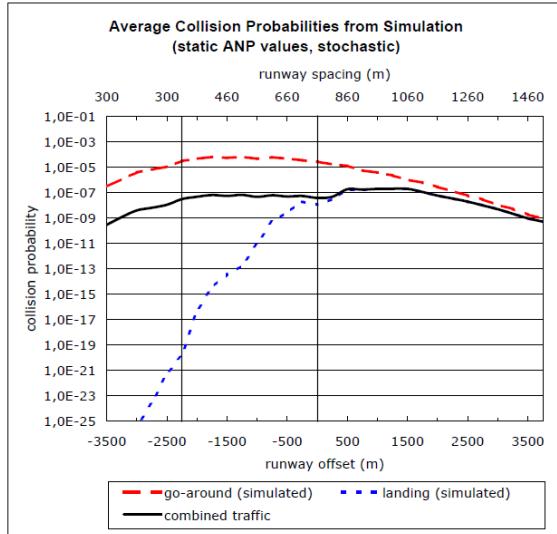


Figure 8. Average simulated collision probabilities

Fig. 9 shows the same results for “dynamic” ANP values obtained from [17] and adaptively assigned to aircraft depending on their flight phase (depending on current instrument landing capabilities, etc.). Since these values are much,

much smaller (a few meters compared to 0.2 NM), reflecting good navigation performance in the TMA, the results differ greatly in magnitude (note the adapted scale). As a secondary result, the dependency on runway staggering increases greatly. The collision risk for landing operations becomes negligible by TLS standards. The shape of the curve indicates that safety hotspots exist, when runways are not staggered at all and when the staggering enables airborne arriving and departing aircraft to interact (>1950 m offset). The operational traffic mix does not reveal any further information (other than a collision risk that barely exceeds $1 \cdot 10^{-20}$ per operation).

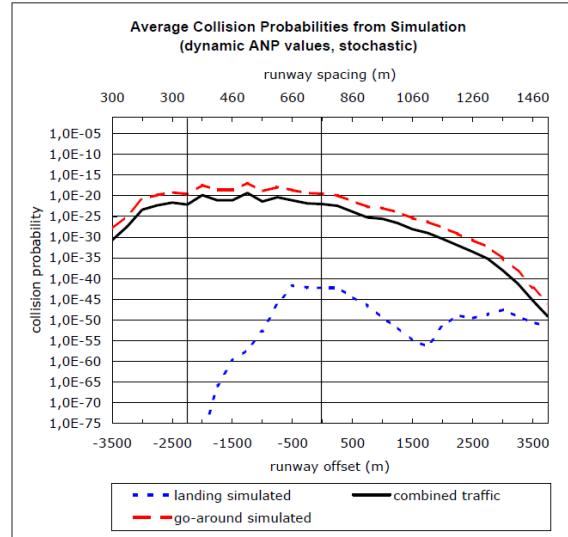


Figure 9. Average simulated collision probabilities

F. Conclusion

Simulative safety assessment with Monte-Carlo style stochastic parameter variation leads to enhanced, potentially more realistic results that indicate a safety-stabilizing impact of the 1/5 runway placement rule. For an operational traffic mix of one missed approach per one thousand landings, the results obtained with fixed ANP values in the range of RNAV precision approach give strong evidence that the average risk per inbound operation becomes somewhat “constant” (with a fluctuation of “just” two decades). Similar results are not as easily produced through geometric analysis and subsequent safety assessment. In particular, the scientific base that supports the 1/5 runway placement rule could not be comprehended through geometric analysis alone. The additional insight gained, greatly helps in understanding the simulative results, though.

The results obtained from simulations with close-to reality ANP values, distinct for each flight phase, show that the current technological collision risk considering state-of-the-art navigation technology lies far lower. The authors conclude that the 1/5 rule relates to earlier navigation aids (the oldest in service need to be considered), that the 1/5 rule might incorporate more complex considerations like weather influence, and that the scientific foundation for its design could likely be flight technical and navigational error modeled in a “static” probabilistic fashion.

IV. OUTLOOK

With this paper, we presented a quantitative safety assessment concept which proves to be applicable to practical problems, i.e. validating existing standards such as those put down in ICAO Annex 14 and PANS-OPS and to provide insight into the contributors of the air traffic system's overall safety. At this point, we have operationalized a technologically objective safety metric that conforms to the so-far promoted target levels of safety specified by ICAO and Eurocontrol. While this paper presented the application to a planning rule for the design of airports as a proof-of-concept, future work shall focus the dynamics of air traffic control, identifying controller strategies for risk mitigation, modeling procedures and human factors, and extensively simulating traffic. The safety metric shall take the role of a target function to evaluate simulated performance.

REFERENCES

- [1] ICAO DOC 9859 AN/460, "Safety Management Manual (SMM)," International Civil Aviation Organization, Montreal, Canada, 2006.
- [2] ICAO, "Annex 10 - Aeronautical Telecommunications, Volume 1," 6th Edition, International Civil Aviation Organization, Montreal, Canada, 2006.
- [3] ICAO DOC 9830, "Advanced Surface Movement Guidance and Control Systems (ASMGCS) Manual," International Civil Aviation Organization, Montreal, Canada 2004.
- [4] ICAO DOC 8168, "Procedures for Air Navigation Services – Aircraft Operations (PANS-OPS)," 5th Edition, International Civil Aviation Organization, Montreal, Canada, 2006.
- [5] CAATS, "Safety Assessment Techniques Database," Nationaal Luchten Ruimtevaartlaboratorium (NLR), Amsterdam, Netherlands, 2005 (from <http://www.caats.isdefe.es/home.html?dissemination>, retrieved: 2010-02-10).
- [6] I. Laudeman, S. Shelden, R. Branstrom, and C. Brasil, "Dynamic Density: An Air Traffic Management Metric," Report NAS/TM-1998-112226. NASA Ames Research Center, Moffett Field, California, USA, 1998.
- [7] A.L. Breitler, M. Lesko, and M. Kirk, "Effects of sector complexity and controller experience on probability of operational errors in air route traffic," Federal Aviation Association, Report DTFA01-95-C-00002, 1996.
- [8] P. Kopardekar, "Dynamic Density – A Review of Proposed Variables, FAA NAS Advanced Concepts," Federal Aviation Association, Report ACT-540, 2000.
- [9] J. Djokic, H. Fricke, M. Schultz, and C. Thiel, "Air traffic Complexity as a Safety Performance Indicator," Modern Safety Technologies in Transportation (MOSATT) Conference 2009, Zlata Idka, Slovakia, 2009.
- [10] R.A. Paielli, and H. Erzberger, Paielli, "Conflict Probability Estimation for Free Flight," Journal of Guidance, Control, and Dynamics, Vol. 20, No. 3, May-June 1997.
- [11] H. Fricke, "Integrated Collision Risk Modelling – CoRIM: An airborne Safety and Protection System for upcoming ATM Concepts," Berlin 2001.
- [12] H. Blom, B. Bakker, M. Everdij, and M. van der Park, "Collision Risk Modeling of Air Traffic," Nationaal Lucht- en Ruimtevaartlaboratorium (NLR), Amsterdam, 2004.
- [13] M. Kietzmann and H. Fricke, „Entwicklung eines quantitativen Bewertungsmodells zur objektiven Beurteilung des Einflusses neuer Betriebsverfahren auf die Sicherheit im Luftverkehr," DGLR Jahrestagung 2004, Dresden, Germany.
- [14] S. Kinnersley and I. Wilson, "INTEGRA Metrics & Methodologies Safety Metrics Technical Definitions", EUROCONTROL, Budapest, Hungary, 2000.
- [15] ICAO DOC 4444 ATM/501, "Procedures for Air Navigation Services – Air Traffic Management (PANS-ATM)," International Civil Aviation Organization, Montreal, Canada, 2001.
- [16] ICAO DOC 9613 AN/937, "Performance-based Navigation Manual (PBN)", 3rd Edition, International Civil Aviation Organization, Montreal, Canada, 2008.
- [17] C. Thiel, „Untersuchung der Navigationsgenauigkeiten von Luftfahrzeugen im Flughafennahbereich – Modellierung und Implementierung dreidimensionaler Verteilungsfunktionen unter Berücksichtigung der Navigationsinfrastruktur an großen Verkehrsflughäfen," TU Berlin, Germany, 2009.
- [18] C. Thiel and H. Fricke, "Collision Risk on final Approach – a radar-data based Evaluation Method to assess Safety," unpublished (to appear at ICRAT 2010).
- [19] ICAO DOC 9643 AN/941, "Manual on Simultaneous Operations on parallel or near-parallel Instrument Runways (SOIR)," International Civil Aviation Organization, Montreal, Canada, 2004.
- [20] ICAO, "Annex 14 – Aerodromes, Volume I, Aerodromes Design and Operations," 5th Edition, International Civil Aviation Organization, Montreal, Canada, July 2009.
- [21] A320 Flight Manual.

Study of SESAR Implied Safety Validation Needs

Jelmer J. Scholte and Henk A.P. Blom

Air Transport Safety Institute,
National Aerospace Laboratory NLR
Amsterdam, the Netherlands
scholte@nlr.nl, blom@nlr.nl

Abstract—Safety validation of changes to an individual organization's local ATM system has become common practice in Europe. However, the SESAR program is planning changes in air traffic operations in Europe that go much further than changes to a local ATM system. This paper identifies the issues on which safety validation approaches need extensions, in order to move from safety validation of changes to a local ATM system to safety validation in SESAR. Subsequently, it identifies approaches that address the identified extension needs. This way an integrated view is developed from the fragmented research results in this area.

Keywords-Safety validation, ATM, SESAR.

I. INTRODUCTION

Safety validation of changes to an individual organization's local Air Traffic Management (ATM) system has become common practice in Europe. As part of this, Air Navigation Service Providers (ANSPs) are required by applicable safety regulations [21],[18] to hand over a positive safety case for regulatory approval prior to introducing a change. However for future changes in ATM, it is highly questionable whether assuring compliance to [21],[18] is effective for SESAR. For example, [21],[18] adopt a conservative approach regarding airborne safety nets: both assume that safety risk reduction by safety nets is taken into account neither in the safety target nor in the safety risk assessment. As a consequence, current regulations may discourage improvements in safety nets [4].

The Single European Sky ATM Research (SESAR) program is planning changes in air traffic operations in Europe that go much further than changes to a local ATM system. SESAR concepts of operations include changes for a multitude of stakeholders including many ANSPs, airlines and airports. The safety of such operations does not only depend on these stakeholders' individual performance, but also on their interactions. Because SESAR strives for ambitious objectives addressing almost contradictory Key Performance Areas (KPAs)¹, the changes to be made are fundamental. This increases even more the importance of addressing safety validation from the concept development start. In early design phases changes to concepts are still relatively easy to make, which makes the provision of feedback to designers the focus of safety validation. Only when the concept of operation matures, the focus of safety validation shifts to derivation of safety requirements and finally confirmation that the concept as developed is indeed safe.

Alberto Pasquini

Deep Blue
Rome, Italy
alberto.pasquini@dblue.it

In this paper issues are identified on which safety validation approaches need extensions, in order to move from safety validation of an ANSP's change to safety validation in SESAR. Subsequently, it is identified which approaches are available to address these extension needs. Although these kinds of questions are being addressed by several researchers inside and outside SESAR, a drawback is that this research is documented in a very fragmented way, which makes it impossible to grasp a complete picture. The aim of this paper is to review these fragmented sources and to provide an integrated view.

The paper is organized as follows. Section II lists relevant studies regarding safety validation needs. Section III discusses safety validation needs identified from SESAR sources. These needs concern two categories: needs regarding organizing safety validation and needs regarding safety assessment. In Sections IV and V approaches are identified that aim to address these two categories of needs. Section VI provides concluding remarks.

II. STUDIES ON SAFETY VALIDATION NEEDS

The methodology widely in use by Air Navigation Service Providers over Europe for safety assessment of changes to their local ATM system is the Air Navigation System Safety Assessment Methodology (SAM) [15]. The current section introduces two series of studies addressing additional validation needs. One series of studies has developed the European Operational Concept Validation Methodology (E-OCVM) [16],[17]. The second series of studies [39]-[48] has been conducted during the SESAR definition phase.

E-OCVM [17] has been developed in order to organize validation from the early concept life-cycle on. E-OCVM provides a common structure to an iterative and incremental approach to operational concept validation, and consists of three elements:

- A Concept Lifecycle Model that reflects the maturity of the concept under investigation (see Fig. 1);
- A Structured Planning Framework that guides planning validation activities; and
- A Case-Based Approach used for providing key stakeholders focused information in an easily understood format.

The main part of this research has been conducted within the European Commission sponsored CAATS II project, and is documented in [35] and [39].

¹ The KPAs for ATM are [29]: access & equity, capacity, cost-effectiveness, efficiency, environment, flexibility, global interoperability, participation by the ATM community, predictability, safety and security.

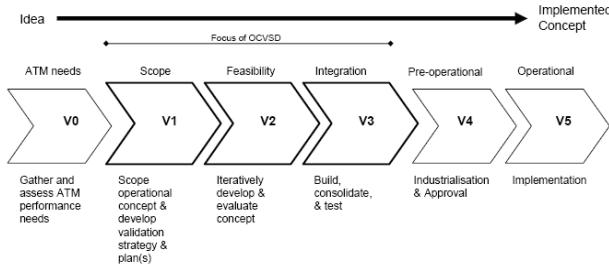


Figure 1. The Concept Lifecycle Model from E-OCVM [17]

The handover by an ANSP of a positive safety case to a regulator as required by [21],[18] effectively applies to phase V4 or V5 of E-OCVM's Concept Lifecycle Model. E-OCVM however poses specific requirements on the outputs of safety validation at the end of each of the Research and Development (R&D) phases (V0, V1, V2, and V3). Effectively, these specific requirements ask for providing feedback to concept developers helpful to reduce the risks associated with new concepts, and structuring evidence into a presentable format that helps stakeholders identify the answers to their key questions.

During the SESAR definition phase, the safety validation needs that emerge for advanced ATM developments have been studied in [40]-[49]. Each of these studies addresses a specific aspect of importance for safety validation in SESAR. [40] provides an overview of the current ATM safety regulatory framework in Europe. [41] summarizes the basic principles of safety regulation, and presents a vision for the future of ATM safety regulation in which issues identified for the current arrangements are addressed. [42] investigates the elements of the SESAR concepts with respect to the impact on and feasibility for safety regulation, and the impact of regulations on the concept elements. [43] studies the impact of SESAR concepts and procedures on safety regulations. [44] describes regulatory and legislative planning including roadmaps for SESAR's 'transversal areas'; these contribute to ensuring that all operational improvements will comply with appropriate safety, security, environment, human performance and contingency requirements and objectives. [45] defines a concept validation methodology that aims to address the complexity of the SESAR concepts. [46] studies the development strategy, including details on aim, content and deliverables in terms of the maturity of the concepts. [47] provides a safety management plan. It aims for an integrated approach to safety related activities, and for establishing an aligned vision for the future of ATM safety that will meet the needs of all stakeholders, now and in the future. [48] describes the system engineering methodology that aims to support SESAR's technical definition in line with the development strategy of [46]. And finally, [49] defines a master plan for management structures and processes.

III. SAFETY VALIDATION NEEDS IDENTIFIED FROM SESAR SOURCES

This section presents safety validation needs that are identified from the SESAR sources presented in Section II.

Table I provides an overview, distinguishing needs regarding organizing safety validation and needs regarding safety assessment.

TABLE I. OVERVIEW OF SAFETY VALIDATION NEEDS AND THEIR SESAR SOURCES

Needs regarding organizing safety validation
<ul style="list-style-type: none"> Addressing E-OCVM requirements [45] Managing relations of the safety case with other cases [44] Addressing the multi-stakeholder nature [42] Addressing future safety regulations [40]
Needs regarding safety assessment
<ul style="list-style-type: none"> Producing a macro safety case [47] Addressing the success side [47] Covering performance of human operators [47] Identifying unknown emergent risks [47] Covering organizational safety [44]

A. Needs regarding organizing safety validation

Addressing E-OCVM requirements: [45] identifies the European Operational Concept Validation Methodology (E-OCVM) [17] as the basis of SESAR's concept validation methodology. Adherence to E-OCVM requirements aims to ensure that stakeholders can make a well-informed decision on further development of a concept, avoiding that the necessity of concept safety improvement is identified in a late stage of development, when modifications are extremely expensive.

Managing relations of the safety case with other cases: SESAR's regulatory and legislative planning document [44] identifies the need for an integrated management approach for all KPAs including safety. Management of performance during design phases is organized via E-OCVM's case-based approach, in which each case focuses on one performance aspect, e.g., safety, business, environment, or human factors.

Cases are usually managed by domain specialists, with the human factors case being managed by human factor specialists, and the safety case by safety analysts. Different domains have different methods and techniques, usually at different levels of consolidation. The result of this partition of work can be a complete separation of the cases, which can affect the efficacy and efficiency of development and validation. There is thus a need to manage relations of the safety case with other cases.

Example results of separation of cases include: 1) real-time simulations focusing on human factor aspects of a concept, without consideration for safety; 2) use of inconsistent assumptions in a safety case and another case, leading to incompatible results and difficulty in interpreting results by decision-makers regarding further concept development.

Addressing the multi-stakeholder nature: [42] identifies that the SESAR operational concept will fundamentally change the roles of many of the ATM stakeholders and, importantly, that these roles will change dynamically within the operation as a flight progresses. This will result in new safety responsibilities and new interfaces between stakeholders. Examples of such changes are in the fields of airspace organization & management, separation provision, and collision avoidance. Necessary precautions should be taken to ensure an appropriate approach towards safety for SESAR in its widest sense. This

includes enabling safe implementation of SESAR concepts, minimizing project risks and related costs, and supporting decision-makers and investors in their requirements to provide information and the discharge of their explicit responsibilities and accountability towards safety in ATM. These conclusions of [42] emphasize the need to properly address the multi-stakeholder-nature of advancing air traffic operations.

An illustration of this need is the development of Continuous Descent Approach (CDA) [33]. The safe execution of CDAs will depend on the roles and collaboration of pilots and air traffic controllers (ATCos). This means that only the reliability of involved interacting technical systems needs to be considered (e.g., via [19]), but also that their overall joint behavior and performance needs to be well analyzed and captured in the CDA development as well as in the safety validation.

Addressing future safety regulations: Even though ATM safety regulations have contributed to the successful delivery of an acceptably safe ATM system across Europe so far, significant issues exist with respect to the current regulatory framework. The main issues that might impact safety of a future ATM system identified by [40] are in the field of:

- Solving the fragmentation and variability in regulations over different domains of air transport, and in the interpretation of regulations over European countries;
- Improving safety accountability: The complex safety regulatory framework and the often detailed and prescriptive nature of safety regulations easily result in confusion over safety accountability;
- Reducing duplication of regulations, as overlap and contradictions lead to confusion and difficulty;
- Reducing complexity of regulation, which otherwise leads to ambiguity regarding compliance; and
- Improving cost effectiveness: it should be clear how ATM safety regulation contributes to cost-effective management of safety.

From this, [40] concludes that developing the ATM safety regulatory framework will be essential to the success of SESAR, and that this improvement should aim to provide a clear, unambiguous set of regulations integrated with the safety regulation of the other parts of the air transport industry. In validation of a concept it should thus be realized that it will eventually need to be proven sufficiently safe according to the safety regulatory framework that will be in force at the time of regulatory approval and implementation of the concept.

B. Needs regarding safety assessment

Producing a macro safety case: SESAR's safety management plan [47] describes that safety assessments in aviation and ATM industry have often focused on individual concept elements, rather than on the joint effect on safety of multiple changes in air traffic operations. SESAR however is defining advanced developments to air traffic operations, consisting of multiple local changes by various stakeholders. The relations and interactions between such individual

operational changes need to be properly assessed. [47] identifies the need for a macro safety case for this, which is to be accompanied by an approach in defining suitable safety targets at an appropriate level for the macro case.

Addressing the success side: Safety assessments in aviation and ATM industry have often focused on what happens if a new or changed system fails in some way, whereas the potential positive contribution of the change is often left unaddressed. Likewise, the positive contribution of SESAR to aviation safety should also be considered, instead of focusing on failures of ATM only. From these observations from SESAR's safety management plan [47] the need to address the success side of a change is identified.

Covering performance of human operators: In future concepts proposed by SESAR, human operators will maintain a central position in ATM. Therefore the safety of air traffic operations will remain dependent on the role of human operators. So far, many safety techniques have not comprehensively covered the role of human operators in the ATM system. [47] emphasizes this need to cover performance of human operators appropriately in safety assessments.

Identifying unknown emergent risks: In [47] it is explained that with the introduction of advanced SESAR concepts yet unknown emergent risk may appear: new behavior and hazards will emerge that have not yet been seen before. Identification of such emergent risk is crucial to be able to take it into account in safety assessment and feedback to design.

Covering organizational safety: SESAR's regulatory and legislative planning document [44] identifies the need for an integrated management approach for safety and other KPAs. The way in which such management system in the eventual operations will be organized can have significant consequences for safety; therefore organizational aspects need to be taken into consideration in safety assessment.

IV. APPROACHES ADDRESSING NEEDS REGARDING ORGANIZING SAFETY VALIDATION

This section presents available approaches for each need identified in Section III.A regarding organizing safety validation.

A. Addressing E-OCVM requirements

E-OCVM [17] poses specific, new requirements to safety assessment, which all boil down to optimal information provision for enabling effective and efficient development and validation processes. Only since recently it has been studied how to tailor safety assessment to the maturity of the concept, and how to satisfy the E-OCVM requirements for its R&D phases V0 through V3. Example plans aiming for E-OCVM compliant safety case development for advanced concepts in these phases are available (e.g., [47] and [23]). Table II shows the CAATS II proposed safety validation activities per E-OCVM R&D phase. However, no publicly available examples of E-OCVM compliant safety cases for advanced concepts in these R&D phases have been identified.

TABLE II. SAFETY VALIDATION ACTIVITIES PER E-OCVM PHASE (BASED ON [39])

R&D phase of E-OCVM	Safety validation activities
V0: ATM needs	Identification of ATM safety performance needs (e.g., safety targets), and support to the identification of ATM barriers that need to be alleviated to reach the ATM needs.
V1: Scope	Safety analysis to determine an appropriate validation strategy, and to provide safety feedback to the development process.
V2: Feasibility	Safety analysis to determine feasibility of the concept, and to provide safety feedback to the development process.
V3: Integration	Safety analysis to provide evidence for the safety of the further detailed concept, and to provide safety feedback to the development process.

B. Managing relations of the safety case with other cases

The risk of thinking for a safety case is that other validation aspects like human factors or business tend to become out of sight for the safety experts, and the other way around with experts of other cases. Moreover, concept designers do not have the luxury to optimize for each separate case. One design should accommodate all cases. Managing relations between cases in the design phases could improve the efficiency of the validation process and increase the synergies between the analyses done by different experts.

Here, only relations in the field of concept evaluation and validation are considered. Relations with the concept development process are explicitly not considered. Concept development is often a struggle to satisfy objectives in several or all KPAs, with the role of validation being primarily in evaluation of concepts. For example, around airports environment and safety are often in conflict: a procedure developed for noise abatement could negatively impact the safety case. Such relations are not considered here, as decision-makers are primarily responsible for balancing different KPAs, and concept developers are primarily responsible for developing concepts in accordance with the objectives. The validation concerns the evaluation of the proposed concept regarding the KPAs.

Within the CAATS II project, a framework for managing relations between cases has been proposed [8]. As depicted in Figure 2, this framework distinguishes relations between case teams, the case generation processes, the cases themselves, and the outputs. Example relations are: 1) different cases provide complementary but coherent outputs; and 2) different cases use the same validation exercises where possible (e.g., simulations or operational trials).

More specifically, the human factors and safety case clearly relate, with a clear overlap of activities. The experience in handling this overlap effectively is rather under-developed.

With the environment case, no clear overlap or input-output relations of the safety case are identified. A scoping issue is which of these two cases should cover third party risk.

Finally, the business case integrates the results from all other cases, including the safety case. Safety gains or losses caused by the introduction of a new concept must be taken into account in the business case. Models for assessing the

economic value of safety gains or losses caused by the new concept are emerging. Also, the eventual cost of a new concept depends on the identification of unsafe elements in the safety case, as these unsafe elements potentially need mitigations or redevelopment, which are costly in time or budget.

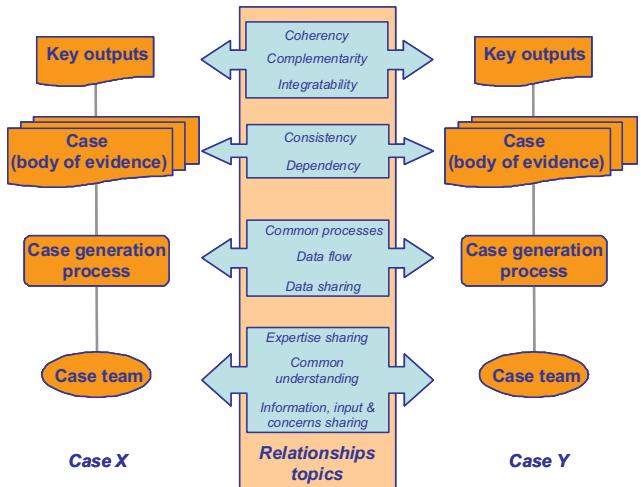


Figure 2. Framework for relations between cases [8]

C. Addressing the multi-stakeholder nature

As advanced concepts will fundamentally change the roles of many of the stakeholders in the ATM system and these roles will change dynamically within the operation as a flight progresses, the multi-stakeholder nature of advancing air traffic operations needs to be addressed.

[23] presents a safety validation framework, which has been developed to incorporate active stakeholder roles during the development and validation of a major change in air transport operations. In its detailed alignment with E-OCVM, the focus during the R&D phases V0 to V3 is on the macro level of institutional conditions, i.e., the interactions between stakeholders' organizations and operational control. Key issue is that during R&D the stakeholders should jointly adopt a goal-oriented approach. This is put in practice via iteration of four processes, in which joint goals are set (set goals), concepts of operations are developed to reach these goals (plan), the consequences for the stakeholders are identified (act), and the concepts are jointly validated (joint safety validation). This is illustrated in Fig. 3. The joint safety validation should ensure that emergent behavior from interactions between the stakeholders is properly addressed.

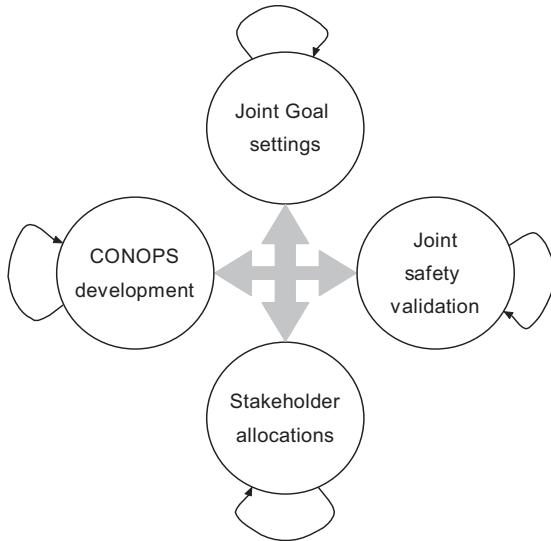


Figure 3. Main processes for active stakeholder involvement in the safety validation framework [23]

D. Addressing future safety regulations

As explained in Section III, significant issues exist with respect to the current regulatory framework, whereas the future regulatory framework needs to be addressed in safety validation. Two complementary approaches are identified for this, which are explained in the following.

Safety fundamentals [51] form a framework of basic safety rules that are independent from the implementation of a design. The main aspects of safety considered in this framework are safety regulation, safety management, operational safety and safety performance. Specific methods are developed to proactively consider operational concepts regarding these aspects early in their development lifecycle. Amongst others, this potentially leads to the identification of needed or anticipated changes in safety regulations, such that these can be properly addressed in concept development and validation. The Safety Screening method [49] has been used for the application of safety fundamentals to early SESAR concepts [42]. Safety Scanning [22] is developed in form of a safety fundamental tool to support authorities in safety regulatory reviews.

[4] has shown that current ATM works with a very large number of minimum separation criteria. The RESET project [38] has verified that in order to accommodate a factor 2 increase in traffic demand over Europe, several of these minimum separation criteria are in need of a significant reduction. Since this cannot be accomplished without conducting a solid safety validation, the aim of RESET is to start the organization of a proper safety validation process for this. Impact assessment of changing minimum separation regulation will make part of this safety validation process.

V. APPROACHES ADDRESSING NEEDS REGARDING SAFETY ASSESSMENT

This section presents available approaches for each need identified in Section III.B regarding safety assessment. Table III provides an overview of identified approaches per safety assessment need.

TABLE III. OVERVIEW OF IDENTIFIED APPROACHES PER SESAR SAFETY ASSESSMENT NEED

SESAR safety assessment need	Identified approaches
Producing a macro safety case	IRP [37] LVNL safety criteria [55] TOPAZ [5]
Addressing the success side	System engineering approach [24] TOPAZ [5]
Covering performance of human operators	Air-Midas [10] CARA [26] Human Assurance Levels [32] Human Factors Case [20] TOPAZ [5]
Identifying unknown emergent risks	Hazard brainstorming [12] Real-time simulations [1] Systemic Modeling [54]
Covering organizational safety	Organizational safety modeling [53] Resilience engineering [28] Scanning on safety fundamentals [51]

A. Approaches towards producing a macro safety case

The need for a macro safety case has a dual character: on the one hand interactions between different operational improvements need to be analyzed on safety, on the other hand suitable safety targets need to be defined for parts of the novel operation. Three approaches ([37], [55], [5]) are identified towards this.

[37] introduces the Integrated Risk Picture (IRP) which aims to integrate safety assessments for individual operational changes, covering their functional interactions and common causes. This provides a top-down approach considering the ATM system as a whole, complementing a bottom-up approach to assess risks associated to hazards affected or newly generated by the introduction of each individual operational change. A ‘baseline’ (IRP 2005) and a future risk picture version (‘predicted IRP’) have been developed. The predicted version models the safety impacts of all known ATM changes, in order to provide an indication whether safety targets can be achieved and to apportion an overall safety target based on the overall ATM contribution to aviation accident risks. The modeled performance of individual ATM elements is used as safety objectives for safety assessments for individual operational changes. The use of IRP is complemented by a ‘Safety Targets Achievement Roadmap’ [56] to interpolate between the baseline and the eventually foreseen situation, taking into account traffic growth and foreseen implementation planning. [24] proposes the use of predicted IRP for SESAR.

[55] presents an approach in developing safety criteria that are based on extrapolating accident rates from the past. The focus is on those accidents that ATC should prevent. This way, all accidents related to separation provision are considered, irrespective of which stakeholder (e.g., ANSP, airline) has causal contributions to the risk. An overall safety target for

ATC-related accidents is apportioned into safety targets on the level of so-called ATC sub-products, which are comparable to parts of a flight forming a logical element within an ATC unit (e.g., taxiing, line-up). [55] proposes that safety assessments consider one or more operational improvements and connect this at the level of the ATC sub-products.

[5] presents the TOPAZ (Traffic Organization and Perturbation AnalyZer) methodology for safety analysis of advanced air traffic operations. It addresses jointly all types of safety issues, including organizational, environmental, human-related and other hazards, and all combinations. Notably, it also considers all stakeholders relevant to the operation in an integrated way, enabling to cover well interactions such as between pilots and ATCos. It makes use of safety relevant scenarios that model the combinatorially many possible interactions between hazards and elements under control by different stakeholders. It features development and subsequent use of a Monte Carlo simulation tool set for selected parts of advanced operations. For other parts and other design options, possibilities are to adopt a qualitative approach, to use sensitivity analysis of the simulations of selected parts, to rerun simulations with adapted parameter settings, and to perform an advanced bias and uncertainty assessment.

B. Approaches in addressing the success side

Whereas safety assessments in aviation and ATM industry have often focused on failures of new systems, there is a need to address the success side of the change. Two approaches ([24], [5]) are identified for this.

[24] presents a system engineering approach to assessing safety. This approach extends upon SAM [15] by adopting the ‘broader approach to safety assessment’ of [25], consisting of complementary success and failure approaches:

- The success approach seeks to show that an ATM system will be acceptably safe in absence of failure;
- The failure approach seeks to show that an ATM system will still be acceptably safe, taking into account the possibility of (infrequent) failure.

This broader approach aims to translate future safety targets that apply to aircraft flights under the operational environment properties of SESAR, to a high level specification of ATM services and their safety objectives. To accomplish this, the broader approach makes use of the predicted IRP [37].

Since its development, the safety assessment methodology TOPAZ [5] has considered success and failure in an integrated way. Hence, it forms a proven approach to covering both the success and failure side of a change. The method uses safety relevant scenarios in which it is modeled how the resolution of hazardous situations depends on the performance of multiple elements, acknowledging that performance variability goes further than the occurrence of failures, and that this plays an important role in safety.

C. Approaches in covering performance of human operators

As safety of air traffic operations will remain dependent on human operators, there is a need to cover their performance

appropriately in safety assessments. Five approaches ([10], [26], [32], [20], [5]) are identified for ATM.

Air Man-machine Integration Design and Analysis System (Air-Midas) [10] is a predictive modeling approach for human operator performance (flight crew, ATC) to evaluate the impact of automation developments in flight management and ATC.

Controller Action Reliability Assessment (CARA) [26] is a human reliability assessment technique, which can be used to quantify human reliability aspects as failure rates and success of mitigation actions in ATM.

[32] explores the use of Human Assurance Levels (HALS), which aim to ensure an appropriate level of Human Factors consideration/ integration in the system design and working practices commensurate with the risk for a particular system function relying on human performance. These HALS are then used at the leaves of fault/ event trees.

EUROCONTROL’s Human Factors Case [20] is a process to systematically manage the identification and treatment of Human Factor issues early in a concept’s lifecycle. In the CAATS II project [7], this Human Factors Case has been formalized for use in R&D in line with E-OCVM. Practices for managing relations between a safety case and a Human Factors case during R&D have already been discussed in Section IV.B.

TOPAZ [5] approach includes a systematic way of incorporating human performance modeling and simulation for ATCos and pilots (e.g., [3]). In [52] this has been extended with a systematic way of modeling the propagation of multi-agent situation awareness differences. In [2] it is explained that while Air-MIDAS is more detailed regarding ATCo and pilot performance, TOPAZ focuses on ATCo and pilot performance impact on accident risk. [11] shows that integration of the two approaches may be of complementary value for both.

D. Approaches in identifying unknown emergent risks

With the introduction of advanced concepts as aimed for by SESAR, unknown emergent risk may appear. Such risk is related to ‘emergent behavior’ which is characterized by what the interaction between multiple local behaviors (both nominal and non-nominal) yields more than the sum of the local behaviors. Three approaches ([12], [1], [54]) are identified.

Hazard brainstorming approaches with experienced pilots and ATCos can be used for identification of emergent risk. HAZID (Hazard Identification, e.g., [9]) aims to identify human failures more effectively by keeping identification separated from hazard analysis and risk mitigation. [12] presents a brainstorming approach that makes use of scenario-thinking rather than application of keywords, using a focus on identification of functionally unimaginable hazards. [13] shows that this can drastically increase the effectiveness over HAZID [9].

Real-time simulations (e.g., [1]) may be used for identification of emergent risk, including risk related to emerging dynamics and interactions of the various elements in foreseen air transport operations. Inserting non-nominal events in the simulations can stimulate this. Risk identification via

expert elicitation can also be improved by involving operational experts in real-time simulations [36].

Recently, there has been a considerable impetus in safety science by approaches for risk assessment by systemic accident models ([27], [28], [31]). Systemic accident models describe the performance of a system as a whole, rather than on the level of events that may go wrong and related cause-effect mechanisms (as fault and event trees). The systemic approach considers accidents as phenomena emergent from variability in performance of interacting entities in an organization. There are four systemic safety risk models in literature for application to ATM: STAMP [31], FRAM [28], IRP [37] and TOPAZ [54]. Stochastic analysis and large scale Monte Carlo simulation of a systemic model, developed with the latter approach, allows filtration of emergent risks from the huge number of less relevant ones. For an active runway crossing operation, [6] compares a safety assessment using stochastic analysis and large scale Monte Carlo simulation with a systemic model versus an event sequence based safety assessment. This showed a significant difference in results due to explicit modeling of the dynamics of the operation, and the concurrent and interacting behaviors of pilots and controllers, which leads to emergent behavior that was neither identified through brainstorming nor through the event sequence based approach [6].

E. Approaches in covering organizational safety

The way in which future ATM will be organized can have significant consequences for safety. Therefore, organizational aspects need to be taken into account in safety assessment. Three approaches ([53], [28], [51]) are identified.

Organizational safety modeling for ATM is being studied in [53], and goes one step further than modeling humans and interactions between multiple humans, in the sense that groups of humans and interactions within and between groups are also considered. In [53], this is done through combined agent- and role-based modeling. The evaluation formalisms used include Bayesian Belief Nets and Monte Carlo simulations.

Resilience engineering [28] acknowledges that safety does not only depend on risk related to breakdown or malfunction, but also on the ability of a system to adjust to current conditions, which continuously change due to the complexity of air traffic operations. Resilience is often reached via a human cognition contribution, e.g., via coordination in unforeseen hazardous situations. Resilience can however also be reached via technological means (e.g., [14]) that help the human in detecting and recovering from latent conditions that undermine the effectiveness of human operators. An example is a tool that helps the operator in detecting hazardous situations resulting from differences in situation awareness.

In Section IV.D scanning on safety fundamentals [51] was discussed as a means towards addressing safety regulations. Other main aspects of safety considered in this framework are safety management, operational safety and safety performance. Consequently, scanning on safety fundamentals (e.g., using [49] or [22]) can be used to pro-actively identify safety management aspects and other organizational aspects of importance for safety.

VI. CONCLUDING REMARKS

This study has shown that several SESAR-identified safety validation needs exist beyond those of ANSPs. These needs appear to be of two categories:

- Needs regarding organizing safety validation, and
- Needs regarding safety assessment.

For each of the identified safety validation needs, relevant approaches have been described. For the needs regarding organizing safety validation, promising approaches are in an early phase of application. For the needs regarding safety assessment, there are multiple approaches (see Table III), of which some have proven to work, and some are new. The experience with the identified approaches is not widely spread. It is recommended to gain experiences with the novel approaches, and to study the complementarity and integration of different approaches in order to combine their strengths.

The expectation is that most of the needs and approaches discussed in this paper also apply to NextGen. Although significant differences exist regarding ATM organization, gaining experience will improve from collaboration between SESAR and NextGen in safety validation.

ACKNOWLEDGMENT

The authors would like to thank Nicolas Fota, Eric Perrin (EUROCONTROL), Marga Martín Sánchez (ISDEF), and Bas van Doorn (NLR) for valuable discussions during the performance of this research as part of the CAATS II project.

REFERENCES

- [1] Antonini, A., and Y. Kermarquer, SAFSIM, simulation for safety insights, EUROCONTROL, 2004.
- [2] Blom H.A.P., K.M. Corker, and S.H. Stroeve, Study on the integration of human performance and accident risk assessment models: Air-MIDAS & TOPAZ, Proc. 6th USA/Europe ATM R&D Seminar, Baltimore, USA, (http://www.atmseminar.org/past-seminars/6th-seminar_baltimore-md-usa-june-2005/papers/paper_098), 2005.
- [3] Blom, H.A.P., J. Daams and H.B. Nijhuis, Human cognition modelling in Air Traffic Management safety assessment, In: Air Transportation Systems Engineering, edited by G.L. Donohue and A.G. Zellweger, Vol. 193 in Progress in Astronautics and Aeronautics, Paul Zarchan, Editor-in-Chief, Chapter 30, pp. 481-511, 2001.
- [4] Blom, H.A.P., M.H.C. Everdij, B.A. van Doorn, D. Bush, and K. Slater, Existing safety assessment methods versus Requirements, RESET D6.1, <http://reset.aena.es/servlet/document.listPublic#>, 2008.
- [5] Blom, H.A.P., S.H. Stroeve, and H.H. de Jong, Safety Risk Assessment by Monte Carlo Simulation of Complex Safety Critical Operations, Eds: F. Redmill and F. Anderson, Proc. 14th Safety critical Systems Symposium, Bristol, UK, February 2006, Springer.
- [6] Blom, H.A.P., S.H. Stroeve, J.J. Scholte, and H.H. de Jong, Accident risk analysis benchmarking Monte Carlo simulation versus event sequences. Proc. ICRAT, Fairfax, VA, June 1-4, 2008, pp. 177-184.
- [7] CAATS II, D17: Guidance document for a typical human factors case, <http://www.caats2.isdefe.es/servlet/document.listPublic>, 2009.
- [8] CAATS II, D28: Guide for a comprehensive incorporation of cases in ATM R&D projects, draft version 0.5, 06-11-2009.
- [9] Civil Aviation Authority, "Hazard analysis of an en-route sector, Volume 1 (main report)", RMC Report R93-81(S), October 1993.
- [10] Corker, K.M., Cognitive Models & Control: Human & System Dynamics in Advanced Airspace Operations, Eds: N. Sarter and R.

- Amalberti, Cognitive Engineering in the Aviation Domain, Lawrence Earlbauem Associates, New Jersey, 2000.
- [11] Corker, K.M., H.A.P. Blom, and S.H. Stroeve, Study on the integration of human performance and accident risk assessment models: Air-MIDAS and TOPAZ, Proc. Int. Seminar on Aviation Psychology, Oklahoma, USA, 18-21, April 2005.
- [12] De Jong, H.H., Guidelines for the identification of hazards; How to make unimaginable hazards imaginable? NLR Contract report 2004-094 for EUROCONTROL (Part of [15]: FHA, Ch. 3, GM B.2), March 2004.
- [13] De Jong, H.H., H.A.P. Blom and S.H. Stroeve, How to identify unimaginable hazards? In: Proc. 25th ISSC, Baltimore, Maryland, 2007.
- [14] Di Benedetto, M.D., A. D'Innocenzo, and A. Petriccone. Automatic Verification of Temporal Properties of Air Traffic Management Procedures Using Hybrid Systems. 7th EUROCONTROL Innovative Workshop & Exhibition. December 2-4, 2008.
- [15] EATMP, Air Navigation System Safety Assessment Methodology (SAM), SAF.ET1.ST03.1000-MAN-01, Ed.2.0, 2004.
- [16] EATMP, "European" Operational Concept Validation Methodology, version 1.0, http://www.eurocontrol.int/valug/public/standard_page/OCVMSupport.html, 2005.
- [17] EATMP, European Operational Concept Validation Methodology, version 2.0, http://www.eurocontrol.int/valug/public/standard_page/OCVMSupport.html, 2007.
- [18] EC, Commission Regulation No 2096/2005 of 20 December 2005 laying down common requirements for the provision of air navigation services.
- [19] EUROCAE, ED78A Guidelines for approval of the provision and use of ATS supported by data communications, December 2000.
- [20] EUROCONTROL EATM, The Human Factors Case: Guidance for Human Factors Integration, version 2.0, 29 June 2007
- [21] EUROCONTROL Safety Regulatory Requirement (ESARR), ESARR 4, Risk assessment and mitigation in ATM, Edition 1.0, 5 April 2001.
- [22] EUROCONTROL, ASRO, Safety fundamentals for safety scanning, Ed.1.0, 3 August 2009.
- [23] Everdij, M.H.C., H.A.P. Blom, J.J. Scholte, J.W. Nollet and B. Kraan, Developing a framework for safety validation of multi-stakeholder changes in air transport operations, Safety Science, Elsevier, Vol. 47, pp. 405-420, March 2009.
- [24] Fowler, D., E. Perrin, and R. Pierce, A systems-engineering approach to assessing the safety of the SESAR Operational Concept 2020 Foresight, 8th USA/Europe ATM R&D Seminar, 2009.
- [25] Fowler, D., G. Le Galo, E. Perrin and S. Thomas, So it's reliable but is it safe?, Proc. 7th USA/Europe ATM R&D Seminar, Barcelona, July 2007, www.atmseminar.org/past-seminars/7th-seminar-barcelona-spain-july-2007/papers/paper_041.
- [26] Gibson, W.H., and B. Kirwan, Application of the CARA HRA Tool to Air Traffic Management Safety Cases, EEC May 2008, http://www.eurocontrol.int/eeu/gallery/content/public/document/eeu/conference/paper/2008/002_Application_of_CARA.pdf.
- [27] Hollnagel E., Barriers and accident prevention, Ashgate, Hampshire, UK, 2004.
- [28] Hollnagel, E., D.D. Woods, and N. Leveson (Eds.), Resilience Engineering – Concepts and Precepts, Ashgate Publishing, 2006.
- [29] ICAO, Global ATM Concept, Doc 9854 AN/458, 2005.
- [30] ICAO, International Standards and Recommended Practices – Air Traffic Services, Annex 11, 13th edition, November 2003.
- [31] Leveson N., A new accident model for engineering safer systems, Safety Science 42:237-270, 2004.
- [32] Mana, P., J.-M. De Rede, and D. Fowler, Assurance levels for ATM elements: Human (HAL), Operational Procedure (PAL), Software (SWAL), 2nd IET International Conference on System Safety 2007 (CP532), p.13–19 doi:10.1049/cp:20070434.
- [33] McMillan, D., The continuous rise of continuous descents, In: International Airport Review, issue 6, 2009.
- [34] Mosquera, D., & G. Cuevas, Separation Minima Standards: Research of Current Applicable Minima Laid Down and Foundations, ICRAT, 2008.
- [35] Pasquini, A., and J.J. Scholte, CAATS II D13, Good practices for safety assessment in R&D projects, in 2 parts, v. 3.6, <http://www.caats2.isdefe.es/servlet/document.listPublic>, October 2009.
- [36] Pasquini, A., S. Pozzi, and G. McAuley, Eliciting Information for Safety Assessment. Safety Science, Vol.46 (10), pp.1469–1482, Elsevier, 2008.
- [37] Perrin, E., B. Kirwan, and R. Stroup, A systemic model of ATM safety: the Integrated Risk Picture, In: Proc. 7th USA/Europe ATM R&D Seminar, Barcelona, July 2007.
- [38] RESET consortium, RESET website, <http://reset.aena.es>.
- [39] Scholte, J.J., H.A.P. Blom, A. Pasquini, and B.A. van Doorn, CAATS II D14 "Guidance document for a typical safety case", v1.91, <http://www.caats2.isdefe.es/servlet/document.listPublic>, October 2009.
- [40] SESAR Consortium, SESAR WP1.6.1/D1, Air Transport Framework The Current Situation, July 2006.
- [41] SESAR Consortium, SESAR WP1.6.1/D2, Air Transport Framework The Performance Target, December 2006.
- [42] SESAR Consortium, SESAR WP1.6.2/ D3, "Air Transport Framework The ATM Target Project, September 2007.
- [43] SESAR Consortium, SESAR WP1.6.2/ D4, Study of Impact of New Concepts and Procedures on Safety Regulations, including Compliance and Synchronisation with ICAO Safety Standards, February 2008.
- [44] SESAR Definition Phase, T3.4.6/D5 Regulatory - Legislative Planning, DLT-0710-346-00-05, version 0.05, May 2008.
- [45] SESAR Definition Phase, Task 4.2.1/D6 Concept Validation Methodology, Part of DLT-0710-421-01-00, version 1.0, May 2008.
- [46] SESAR Definition Phase, Task 4.2.1/D6 Development Strategy, Part of DLT-0710-421-01-00, version 1.0, May 2008.
- [47] SESAR Definition Phase, Task 4.2.1/D6 Sesar Safety Management Plan, Part of DLT-0710-421-01-00, version 1.0, May 2008.
- [48] SESAR Definition Phase, Task 4.2.1/D6 System Engineering Methodology, Part of DLT-0710-421-01-00, version 1.0, May 2008.
- [49] SESAR Definition Phase, WP4.1.12/D5-D6 SESAR Definition Phase, DLT-0701-041-00-081, April 2008.
- [50] Straeter, O., M. Everdij, J. Smeltink, J. Nollet, J. Kovarova, H. Korteweg, A. Burrage, Safety Screening – Experiences in applying a proactive approach to concept development within SESAR, Proc. EUROCONTROL Safety R&D Seminar, Rome, Italy, 24-26 Oct. 2007.
- [51] Straeter, O., Managing Safety Proactively – Experiences on the Implementation of the Safety Agenda at EUROCONTROL, In: Proc. PSAM 8, New Orleans, Louisiana, USA, 2006.
- [52] Stroeve S.H., H.A.P. Blom, and M.N.J. Van der Park. Multi-agent situation awareness error evolution in accident risk modeling. Proc. 5th USA/Europe ATM R&D Seminar, Budapest, Hungary, (http://www.atmseminar.org/past-seminars/5th-seminar-budapest-hungary-june-2003/papers/paper_067), 2003.
- [53] Stroeve, S.H., A. Sharapanskykh, R.M. van Lambalgen, and B. Kirwan, Safety culture analysis by agent-based organizational modeling, In: Proc. 7th EUROCONTROL Innovative Workshop & Exhibition, 2008.
- [54] Stroeve, S.H., H.A.P. Blom, and G.J. Bakker, Systemic accident risk assessment in air traffic by Monte Carlo simulation, Safety Science, Vol. 47 (2009), pp. 238-249 (<http://dx.doi.org/10.1016/j.ssci.2008.04.003>).
- [55] Van den Bos, J.C., H.H. de Jong, and R.B.H.J. Jansen, Apportioned ATC Safety Criteria Based on Accident Rates, In: ATC Quarterly, vol.17 no.3, 2009.
- [56] Vernon G., and E. Perrin, Methodology report for a Safety Target Achievement Roadmap (STAR), EUROCONTROL report, May 2007.

Collision Risk on Final Approach – A Radar Data Based Evaluation Method to Assess Safety

ANP Based Obstacle Assessment Surfaces

Christoph Thiel and Hartmut Fricke

Chair of Air Transport Technology and Logistic
 Technische Universität Dresden, Germany
 thiel@ifl.tu-dresden.de; fricke@ifl.tu-dresden.de

Abstract— Many major airports around the world are facing the problem of highly congested airspace and are therefore seeking ways to enhance capacity. Innovative RNP/ RNAV procedures in terminal areas, in particular RNP/ RNAV procedures for the final approach segment may be a possible solution due to increased flexibility when using the available airspace. However, these procedures must be designed according to their navigational performance requirements to ensure safe operations. Measuring safety of upcoming RNAV approach procedures in terms of navigational accuracy is crucial for their implementation at airports, as there is a need to develop specific obstacle assessment surfaces (OAS) and collision risk models (CRM). Designing specific OAS is essential for future airport development if benefits of improved navigational performance shall be fully exploited. This paper presents a method to determine actual navigational performance (ANP) during the final approach phase and a strategy for calculating ANP-based OAS executed here for an ILS final approach by means of radar data evaluation. Radar data will be used for statistical analysis of approach path deviations during the final approach phase and for modeling specific OAS based on the derived deviations.

Safety; Collision probability; Actual navigation performance (ANP); Obstacle Assessment Surfaces; Radar data analysis

I. INTRODUCTION

The objectives of the SESAR ATM Master Plan [1] are amongst others to enhance the capacity of the European air traffic system by a factor of three and at the same time enhance safety by a factor of 10. To reach these ambitious goals, new flight procedures need to be developed and implemented, especially at already nowadays highly congested terminal areas (TMA) around major European airports. Take off and landing, which take place here, are still the most critical flight phases, as the majority of all accidents (56% of all fatal accidents) occur during these phases, although they only cover about 6 % of the total flight time for a typical 1.5 hour flight. When only considering final approach and landing, the ratio is at 4 % flight time to 36 % of fatal accident likelihood [2].

So the flight phases take off and landing are preferred fields to improve safety in the European air traffic system. Improvements may be reached by introducing new approach procedures such as RNP/RNAV approaches, requiring a navigational accuracy during final approach (non-precision) of

at least 0.3 NM. In detail, this expresses a Required Navigation Performance (RNP) of 0.3 NM cross and along the desired flight track for 95% of the flight time. This equals the two sigma interval of a normal distribution. RNAV approaches may increase safety, which could not be analytically demonstrated until now, but doubtlessly it will improve capacity of airports, especially in obstacle rich environments, as the lateral and vertical approach path is more flexible to adapt to environmental requirements.

However, measuring safety in terms of navigational accuracy of upcoming RNAV approach procedures is crucial for their implementation at airports, as e.g. there is a need to develop specific obstacle assessment surfaces (OAS) and collision risk models (CRM) for these kind of procedures according to ICAO airspace design requirements. Along those, ICAO PANS-OPS [3] methods to construct OAS and apply CRM are described already for e.g. innovative GBAS CAT I approach procedures. However, they are not specifically explored - they simply use the same OAS and CRM calculation method as for the reference ILS CAT I approach, with only few adjustments for OAS constants.

This paper presents a method to determine the navigational accuracy (actual navigation performance – ANP) which shall always be higher than the design RNP value. We focus on the final approach segment based on a radar data analysis taken from live traffic in 2008. Due to the lack of radar data of an innovative RNAV approach procedure, the analysis is based on an ILS final approach segment. Nevertheless, the described method can be fully adopted for RNAV final approach procedures when such data will become available, although results may be different from those shown here, due to different navigational performance. Knowing navigational accuracy of specific flight segments is essential to estimate collision risks [4], [5] and to derive obstacle assessment surfaces for a specific procedure. Therefore finally a method will be shown, how to construct OAS using determined ANP for the investigated procedure.

II. NAVIGATIONAL ERRORS

Collision risk during final approach firstly depends on the relative position of an object (obstacle) to the nominal approach path and secondly on the navigational accuracy of

approaching aircraft, so their along and cross track tolerance. The following section will give a systematic overview of navigational errors during final approach depending on the flown procedure.

A. Navigational errors during conventional approach procedures

The navigational accuracy during conventional approach procedures, of which the ILS approach is the most important one, is influenced by the following three error categories:

- Errors of ILS ground equipment – localizer (LLZ) and glide path (GP) antenna signals (GEE, ground equipment error),
- Errors of airborne equipment – LLZ and GP receiver and indicator (AEE, airborne equipment error) and
- The difference between desired and true trajectory – errors induced by the pilot/ autopilot (FTE, flight technical error)

The Total System Error (TSE) is denoted as the root sum square of these three categories:

$$TSE = \sqrt{GEE^2 + AEE^2 + FTE^2}$$

B. Navigational error of RNP-RNAV procedures

The inability to achieve the required navigation performance during RNAV procedures may be due to navigation errors related to aircraft tracking and positioning in the context of on-board performance monitoring and alerting as it is mandatory for RNAV procedures. According to ICAO Manual on Performance Based Navigation (PBN) [6] the navigational errors for RNAV contain the following three main error categories:

- Path definition error (PDE)
- Flight technical error (FTE) and
- Navigation system error (NSE)

The PDE occurs when the path defined in the RNAV system database does not correspond with the desired path. Flight Technical Errors are again errors induced by the pilot/autopilot including display errors. The NSE refers to the difference between the aircraft's estimated position and actual position. So this is the error of the multi-sensor navigation system, as e.g. the error of the GPS. The TSE is again the root sum square of three error categories, here PDE, FTE and NSE.

C. Accuracy requirements for current RNAV approach procedures

The ICAO PBN Manual [6] defines two types of RNAV approach procedures which are applicable to the final approach segment. First, the "non precision alike" RNP APCH which is defined as an RNP approach procedure that requires a lateral TSE (Along Track and Cross Track) of ± 1 NM in the initial, intermediate and missed approach segments and a lateral TSE of ± 0.3 NM in the final approach segment. Second, the RNP AR APCH (authorization required), which is defined as RNP

approach procedure requiring a lateral TSE of at least ± 0.3 NM and down to ± 0.1 NM for all approach segments.

D. Accuracy requirements of upcoming RNAV approach procedures

With the application of RNP concepts to approach procedures, and in particular to upcoming precision approaches, the All Weather Operations Panel (AWOP) had concern in also addressing a required vertical navigational accuracy beside along and cross track tolerances. As a result, a range of RNP types were defined from RNP 0.3 down to RNP 0.003/z, where z reflects the requirement for vertical guidance. The following Tab. I collects all proposed RNP types with vertical and lateral TSE according to [7]:

TABLE I. PROPOSED RNP TYPES FOR FINAL APPROACH SEGMENTS [7]

RNP Type	Required Accuracy (95% containment)	Description
0.003/z	± 0.003 NM [$\pm z$ ft]	Planned for CAT III Precision Approach and Landing including touchdown, landing roll and take-off roll requirements. (ILS, MLS and GBAS)
0.01/15	± 0.01 NM [± 15 ft]	Proposed for CAT II Precision Approach to 100 ft DH (ILS, MLS and GBAS)
0.02/40	± 0.02 NM [± 40 ft]	Proposed for CAT I Precision Approach to 200 ft DH (ILS, MLS and GBAS)
0.03/50	± 0.03 NM [± 50 ft]	Proposed for RNAV/VNAV Approaches using SBAS or GBAS
0.3/125	± 0.3 NM [± 125 ft]	Proposed for RNAV/VNAV Approaches using Barometric inputs or SBAS inputs.

E. Errors of radar antenna system

Before applying the method to analyse the aircraft's ANP by means of radar data analysis one more additional error needs to be considered.

Although the TSE of the researched procedure (regardless of which) can be measured with the here applied radar data analysis, it should be noticed, that this measurement may be non-precise due to erroneous radar antenna system data itself. Due to angular signal characteristic of the radar system, the error will increase with increasing distance to the radar facility. The magnitude of this radar equipment error is not quantifiable right now, due to lack of adequate data. Nevertheless, when focussing on a specific, relative small investigation area, as it is performed here, the error for all covert data points will be within the same range and therefore will have only a small impact for the statistical analysis applied here. Consequently, it could be assumed, that due to this additional error all here shown results has to be considered as conservative results. The TSE during final approach may be effectively lower than the measured one by means of a radar data analysis.

III. DATA SOURCE

The radar data used for the analysis are taken from the Flight Track and Aircraft Noise Monitoring System (FANOMOS) of the German ANSP *Deutsche Flugsicherung GmbH* (DFS). These tracks compile all transponder equipped arrivals and departures at a major German airport for a period of six months (busiest six month – May to October). As these radar data are classified as confidential, the airport has to be treated anonymous and will be named as *investigation airport* in this paper. The recorded data comprehends the aircraft flight tracks in terms of single data points (position of aircraft, identification and aircraft reported altitude via SSR Mode A/C resp. S) recorded at an update rate of 4 sec and complemented by flight plan data. In detail, the following information are forming a data bloc:

- Flight plan data (e.g. date, aircraft type, call sign, used runway, time of arrival/ departure)
- Radar data:
 - Time stamp, measured in seconds from first recording
 - Horizontal positional information – X, Y coordinates in UTM WGS84 format
 - Vertical positional information – altitude in meter above mean sea level
 - Ground speed in meter per second
 - Distance in meter, measured from first recording

Altogether 81'084 flight movements were sampled at the investigation airport and recorded by FANOMOS in this half year time period in 2008. So, the recorded data consist of more the 8.3 million single data points at a resolution of 4 seconds. The data are given in tabular form in ASCII Format.

IV. STATISTICAL ANALYSIS OF APPROACH PATH DEVIATIONS DURING ILS FINAL APPROACH

A. Methodological overview

Based on the available radar data the deviation probability from the nominal flight path will be calculated, performing the following steps:

1. Selection of usable data (ILS approach only)
2. Definition of final approach segment and splitting into investigation segments (cross sections) along the approach path in fixed steps
3. Determination of lateral (cross to the flight path) and vertical deviations from final approach path at every cross section
4. Determination of mean and standard deviation in lateral and vertical direction
5. Modelling of approach path deviations as distributions related to the distance to the threshold
6. Verification of the resulting distribution functions.

B. Selection of useable data

As explained in the previous Chapter III, used radar data conclude all air traffic movements at the investigation airport within a time period of half a year. So, the first step in determining the approach path deviation is to select a specific number of flight tracks out of radar data base, which are doubtlessly precision approaches, as ILS final approach is focused procedure here. Therefore, the following steps were performed to filter the data base:

- Identify the flight phase: Either approach or departure
- Identify the landing direction
- Split precision and non precision approaches

After these steps, altogether 14'500 precision approach flight tracks were found for further statistical analysis. The following Fig. 1 shows an extract of these flight tracks (approx. 700 approaches are shown):

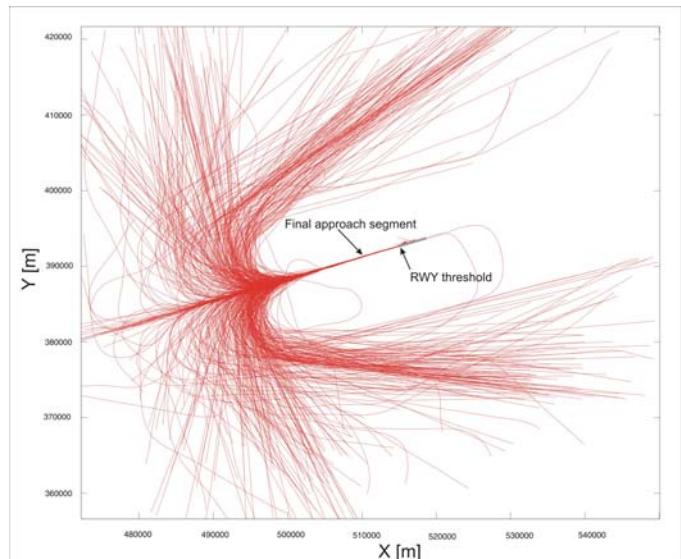


Figure 1. flight track snapshot – ILS approach

Already the visual comparison shows the much higher navigational performance during final approach than e.g. during the intermediate approach phase (concentration of flight tracks on the final approach segment) inline with the presented ICAO RNP/RNAV concept, as presented in Chapter II.

C. Quantification of path deviations

Starting from the landing threshold and following the approach path in the opposite direction, cross section windows were defined at 1000m distance intervals along the approach path ending at the final approach fix (typically located about 8 to 12 NM threshold distance).

Afterwards, the intersection points of the flight tracks at each defined cross section window were determined. This is based on linear interpolation between two radar data points from the flight track information (as update frequency of radar antenna is 0.25Hz distance between two data points ranging between about 150 m and 400 m depending on the final

approach speed). The lateral as well as vertical deviation (altitude deviation) from the nominal flight path was determined for all cross section windows. Subsequently all outliers were removed from the data base by the Grubbs-outlier test. Outliers are data points, which are significantly different from the all other data points. Such outliers were caused by e.g. missed approaches or late LLZ or GP intercepts. In vertical direction only outliers above the glide path were removed, all potential outliers below the glide path were not eliminated from the database, as these data points are important for modelling OAS. Overall, only very far outliers were removed and therefore only very few outliers (less than 0.5%) has to be eliminated.

The following Fig. 2 exemplarily represents the determined flight intersection points for the 6'000m cross section window (6'000 m threshold distance) for roughly 14'500 data points:

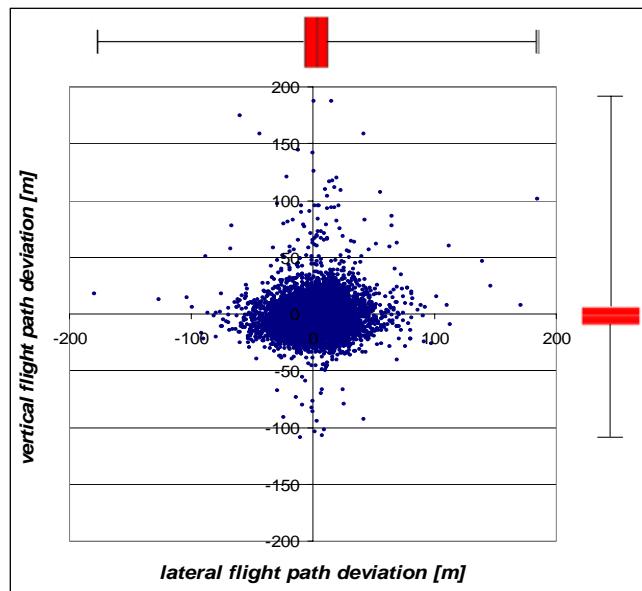


Figure 2. Measured Deviations from nominal track at 6000m distance to threshold

As expected, Fig. 2 shows a clear concentration around the nominal approach path: for the shown cross section window most of measured intersection points are deviating less than 50 m in lateral and less than 25 m in vertical direction. This can also be seen when looking at the Box-Whisker Plots, showing the 50%-quantile (red box), which is less than ± 10 m in range for both directions. For comparison purposes: one dot deviation on the HSI/ PFD for ILS LLZ resp. GP equals to ± 83 m in lateral and ± 28 m in vertical direction for this threshold distance.

D. Quantification of underlying distribution function

Next step in determining the approach path deviations by modelling location probability functions (PDF) is to find a PDF that fits the measured deviations. Air traffic safety research at TU Dresden shows that a normal distribution function is most often best fit for describing navigational accuracy for many approach and departure procedures [8], [9]. Therefore a normal distributed behaviour will be assumed as underlying

distribution. In order to check legitimacy of the assumption of a normal distribution, exemplary all available data of the 6000m section were statistically analyzed. To that purpose, the data will be arranged into classes of a specific number and range using statistical methods. The number of classes and the dimension of their range can be defined freely, but the number of classes should be in between five and about twenty. A common used approximation of the number of required classes is:

$$k \leq 5 \cdot \lg(N) \quad (1)$$

where N corresponds to the sample size (here number of given arrivals). Then the associated class range k_b is calculated from the bandwidth of the ascertainment data:

$$k_b \approx \frac{X_{\max} - X_{\min}}{k} \quad (2)$$

According to this calculation, a number of 21 classes is produced with a class range of 8 m for the lateral direction and respectively 21 classes with a class range of 6 m for the vertical direction. The following Fig. 3 shows exemplarily for the 6000m section, the class frequency according to the determined grading for the lateral and the vertical direction as well as the progress of a normal distribution function:

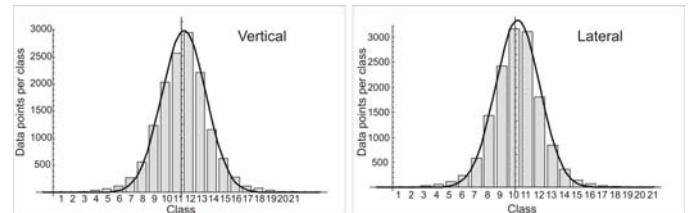


Figure 3. Distribution in lateral direction and approximated normal distribution, 6000m section

Like Fig. 3 shows, the distributions are symmetric and seem to follow normal distributions in lateral as well as in vertical direction. This distribution behaviour was observed for all cross section windows, instead of the cross section which is located directly at the threshold. Here a slightly asymmetric behaviour in the vertical direction was found, which may be due to the beginning of the flare phase. Furthermore in this approach phase aircraft are not longer on instrument approach, but on visual approach. Nevertheless the normal distribution will be assumed for all cross section windows as underlying distribution function.

The normal distribution is depending on the two parameters mean value μ and standard deviation σ , only. The one-dimensional density function of the normal distribution appears as follows:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (3)$$

Now it is possible to determine the mean value as well as the standard deviation in lateral and vertical direction for all 13 cross section windows out of these data points. With assumption of a normal distribution, location probability

functions can be calculated on the basis of these two statistical parameters.

For the cross section window, which is demonstrated exemplarily, the following mean was calculated for the lateral and vertical direction:

$$\mu_{\text{lateral}} = 1.34 \text{ m} \text{ and } \mu_{\text{vertical}} = 2.65 \text{ m.}$$

For the standard deviation follows:

$$\sigma_{\text{lateral}} = 16.25 \text{ m} \text{ and } \sigma_{\text{vertical}} = 11.36 \text{ m.}$$

Like also Fig. 3 shows,, the progresses are following the normal distribution “very well”. The statistical test using coefficient of determination and F-Test documents this objectively too, wherewith the legitimacy of the figure with the distribution of the approach accuracy in lateral and vertical direction is detected through a normal distribution. The Chi-Square-Test, which is also often used for such tests, is not applicable for such big sample sizes.

E. Estimation of distribution parameters for all cross section windows

Now the estimation of the distribution parameters mean and standard deviation are performed for all defined, previous cross section windows along the approach flight path. Furthermore for all cross section windows the statistical test, using coefficient of determination and F-Test, was performed to prove the assumption of a normal distribution. The following Tab. II represents estimated values of standard deviation and mean value for each of the 13 sections in lateral and vertical direction:

TABLE II. DEVIATION PARAMETER DURING ILS APPROACH, ALL SECTIONS, VERTICAL AND LATERAL

Cross section window	threshold distance [m]	Lateral		Vertical	
		Mean [m]	Standard deviation [m]	Mean [m]	Standard deviation [m]
Section 0	0	1.357	9.283	1.236	8.828
Section 1	1000	12.917	12.652	1.265	8.878
Section 2	2000	26.654	16.025	1.690	8.431
Section 3	3000	32.632	18.074	2.779	9.115
Section 4	4000	28.339	17.295	3.653	9.641
Section 5	5000	15.240	16.255	3.569	10.408
Section 6	6000	1.342	16.247	2.649	11.360
Section 7	7000	-9.285	17.521	1.809	12.602
Section 8	8000	-15.575	20.292	1.825	14.026
Section 9	9000	-18.452	23.600	2.192	15.702
Section 10	10000	-20.028	27.542	1.461	17.775
Section 11	11000	-20.774	32.535	-1.234	20.085
Section 12	12000	-21.572	40.349	-5.354	22.313

As expected, Tab. II shows a distance-depending distribution characteristic relating to threshold distance: The greater the distance to the threshold, the greater is the distribution (standard deviation) in lateral as well as vertical direction. Furthermore, in lateral direction, a slightly oscillating behaviour around the nominal flight path could be observed. In vertical direction, for all sections (instead of far threshold distances) the mean value for vertical direction is above the flight path. It's assumed that many hand flown approaches are

intentionally performed slightly above the glide slope for safety purposes.

Now, the distance dependency can be approximated via linear interpolation, wherewith it is possible to determine the given distributions for any threshold distance, analytically.

According to this, for lateral deviation applies the following distance dependency for the standard deviation linear fitted:

$$\sigma_{\text{lateral}}(x) = 0.002x + 8.86 \quad (4)$$

with x = threshold distance in [m] and $\sigma(x)$ = lateral standard deviation in [m]. Thus the standard deviation amounts to 8.86m on threshold and is increasing about 2m per each 1000m distance to the threshold (respective 0.11°). Analogue, for the vertical deviation applies the following distance dependency for the standard deviation:

$$\sigma_{\text{vertical}}(x) = 0.00113x + 6.25 \quad (5)$$

Thus, the standard deviation amounts to 6.25m on the threshold and is increasing about 1.13m per each 1000m threshold distance (respective 0.065°). Both fits (lateral and vertical) are valid for a threshold distance up to 12'000 m. The acceptance was checked again via coefficient of determination and F-Test for both function approximations.

In conclusion, the assumed distance dependency of flight path deviations during ILS final approach – increasing lateral and vertical deviation with increasing threshold distance – could be shown with the presented statistical analysis.

F. Discussion of results

The given deviations can be converted into ANP values according to RNAV convention (95% or two sigma containment in NM) for comparison purposes. The following Tab. III represents ANP values in lateral (cross track tolerance – XTT) and vertical direction (vertical track tolerance – VTT) for some exemplary threshold distances:

TABLE III. EXEMPLARY ANP VALUES (XTT AND VTT) FOR ILS FINAL APPROACH

Distance to threshold	XTT	VTT
0 m	0.012 NM	0.008 NM
2500 m	0.015 NM	0.010 NM
5000 m	0.020 NM	0.013 NM
10000 m	0.031 NM	0.019 NM

As seen in Tab. III, XTT values range between 0.01 NM and 0.03 NM and VTT values between 0.01 and 0.02 NM. Compared with Tab. I, these values are in a range of about CAT I to CAT II approaches. RNP values for CAT III compliance are not reached, although the investigation airport is ILS CAT IIIb equipped. This is most likely due to the investigated time period during summer month. As CAT II/III conditions are very rare (at a guess less than 1% of all approaches are performed under CATII/ III conditions) most approaches were on a CAT I approach. Furthermore the above

given XTT and VTT can now be compared with ICAO Annex 10 [10] requirements for ILS CAT I. The following Fig. 4 represent the comparison of the here estimated lateral deviations (XTT) with ICAO tolerances for localizer (LLZ) signals for an ILS CAT I approach, for both the (\pm) 95% containments are shown:

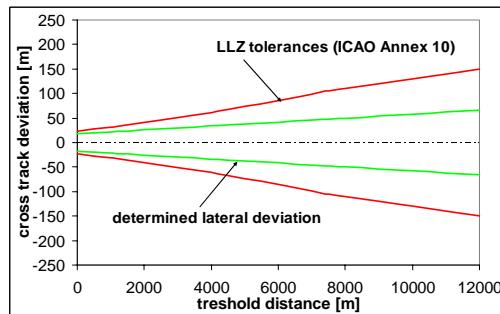


Figure 4. lateral deviations vs. ICAO tolerances for LLZ

Fig. 5 below represent the comparison of the previously estimated vertical deviations (VTT) with ICAO Annex 10 tolerances for glide path (GP) signals. For both, the (\pm) two sigma intervals are shown:

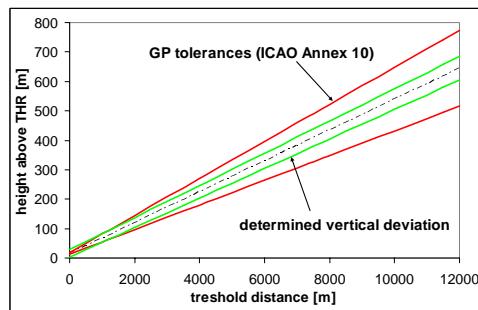


Figure 5. vertical deviations vs. ICAO tolerances for GS

As seen in Fig. 4 and Fig. 5 the derived lateral and vertical deviations are significantly below ICAO Annex 10 requirements, even though derived deviations are TSE and ICAO requirements need to be considered as just GEE, which is only one part of the TSE (see Chapter II). It's assumed that ICAO requirements were developed decades ago on basis of the achievable navigation performance of that time (e.g. ICAO collision risk model [11] was developed in the 70ies decade). But navigation performance is supposed to have significantly improved since then due to technical improvements on ground equipment as well as airborne equipment, which might be a reason for these findings. Higher vertical deviations very close to threshold (see Fig. 5 – less then app. 1000 m threshold distance) may be caused by radar equipment error, due to radar reflecting characteristics for ground near targets.

V. MODELLING OF ANP- BASED OBSTACLE ASSESSMENT SURFACES

A. PANS-OPS OAS approach surfaces

Obstacle assessment surfaces (OAS) according to PANS-OPS [3] are imaginary surfaces which guarantee obstacle free

approach (in detail a collision risk below the Target Level of Safety of 1×10^{-7} per approach), when operating under instrument flight rules (IFR) on precision or non-precision instrument approach. The OAS system is based on collision risk calculation of ILS Collision Risk Model (CRM) [11]. For the CRM collision risk functions (normal distributed PDF in vertical and lateral direction) the 1×10^{-7} per approach probability curve (contour of same collision risk during the precision segment of an ILS approach) is used for surface modelling. At this curve tangential surfaces in trapezoid shape were fitted, to get defined, plane surfaces which guarantee a collision risk less then 1×10^{-7} per approach. The following Fig. 6 shows the 1×10^{-7} per approach probability curve and the tangential placed surfaces, which form the OAS final approach funnel according to [12]:

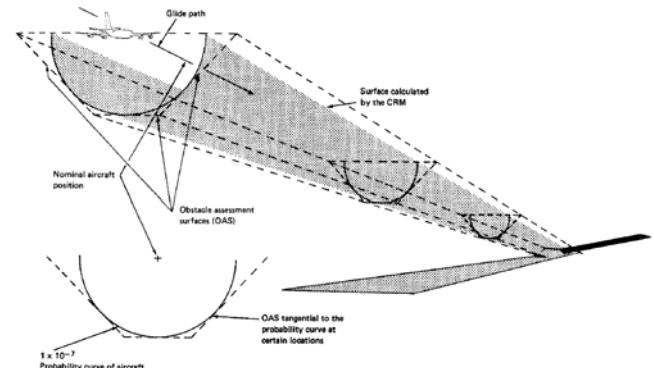


Figure 6. OAS approach funnel according to [12]

As seen in Fig. 6, the approach surfaces are getting closer, the closer the distance to the landing threshold is, so this also shows the angular signal characteristic of the ILS.

B. Potential shapes of ANP based OAS

As seen in the previous chapter PANS OPS defines plane surfaces embedding the CRM 1×10^{-7} collision risk contour. The construction of such surfaces is per se not necessary, as the CRM contour already gives an area of maximum allowable collision risk, but in shape of tapering ellipses that are difficult to describe. Due to simplification matters for airport procedure designers this is approximated conservatively by the described tangential surfaces. But this simplification leads to an overestimation of collision risk in some specific areas, as the surfaces are bigger than the ellipses. By construction of other shapes than the trapezoid surfaces this overestimation could be decreased, but this is always associated with a more complex surface design. The following Fig. 7 shows some potential tangential surface shapes ordered from simple to complex:

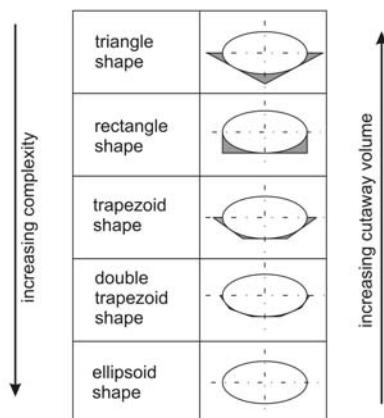


Figure 7. potential surface shapes

With increasing complexity the cutaway volume (volume between ellipse and tangential surfaces) will decrease significantly and additionally the overestimation of collision risk decreases. This cutaway volume can be expressed as overestimation volume in percent of the half-ellipsoid volume. This is for triangle and rectangle shape about 27.3%, for trapezoid shape about 10.3 %, for double trapezoid shape about 3.4% and for the ellipsoid shape 0%. So when only looking at cutaway volumes, the best solution appears to be the ellipsoid shape, but when taking design criteria into consideration this would be the most unpractical solution, as the more complex the shapes – the more complex is the calculation scheme, which makes it unsuitable for airport procedure designers.

So it could be assumed, that the trapezoid shape is the best compromise between complexity and risk overestimation. Therefore the following chapter will describe the method of shape modelling focussing on trapezoid shape. Some more potential shapes of OAS surfaces can also be found in [14].

C. Definition of TLS

As described in subsection V.A, the TLS for ILS operations was set by the ICAO to 1×10^{-7} per approach. But when taking ambitious SESAR goals into consideration, which challenges an increase in safety by a factor of 10, this TLS may be adjusted for future operations. Therefore, future TLS for upcoming final approach procedures may be set to 1×10^{-8} per approach following the SESAR goals.

Consequently, the OAS construction described in the following chapter will be performed using TLS of 1×10^{-7} per approach. Afterwards the impacts of using a 10-times lower TLS of 1×10^{-8} per approach will be shortly discussed according to the derived results.

D. Construction of radar-fitted approach surface

Based on the in subsection IV.E shown deviation functions in lateral and vertical direction, we are now able to construct obstacle assessment surfaces for derived ANP during final approach. Firstly, we need to calculate the size of the 1×10^{-7} contour, in other words to calculate the $1-1 \times 10^{-7}$ quantile of the lateral and vertical PDF. Furthermore the maximum size of an approaching aircraft needs to be considered, here according to

ICAO Annex 14 [13] a category F aircraft was considered, which is e.g. the Airbus A380 as current largest commercial aircraft. The semi-span (40 m) was added to the 1×10^{-7} contour of the lateral distribution function and the distance between the glide path antenna and the lowest point of the landing gear (8 m) was added to 1×10^{-7} contour of the vertical distribution function. The following Fig. 8 shows the general method of determination of the radar-fitted OAS approach surfaces exemplarily for 1'000 m threshold distance:

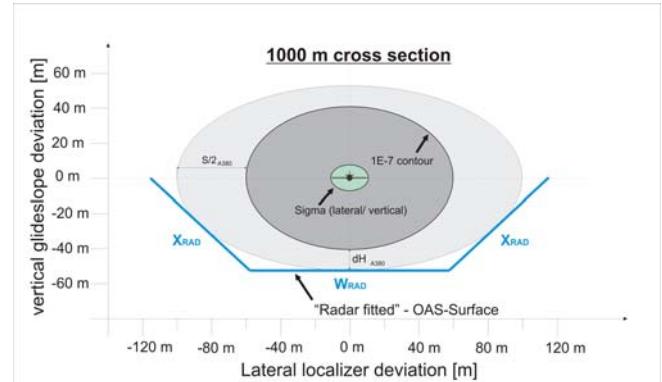


Figure 8. exemplary surface modelling for 1.000 m cross section

The inner (green) ellipse in Fig. 8 shows the dimension of the one sigma interval (standard deviation) for the 1000 m cross section window (see also Tab. II) centred on the nominal flight path. The surrounding darker gray ellipse shows the 1×10^{-7} per approach contour. Moreover, the surrounding light-grey ellipse takes the size of Cat F aircraft into consideration.

On the outer edges of this ellipse tangential surfaces in trapezoid shape (blue lines in Fig. 8) analogue to OAS approach surface were modelled. The tangential surfaces were constructed in such a way, that the volume between the surfaces and the outer ellipse was minimized, in order to have as less refuse as possible. When applying this method for every cross section window, a linear increasing approach funnel around the nominal approach path will be formed, due to the distance dependency of the modelled PDF (increasing distribution with increasing threshold distance) and linearization of distribution parameters. A surface modelling above the nominal flight path is not necessary, as it is assumed that any path deviation above the glide slope signal is uncritical for obstacle assessment. The following Fig. 9 give a top view of the assessed OAS:

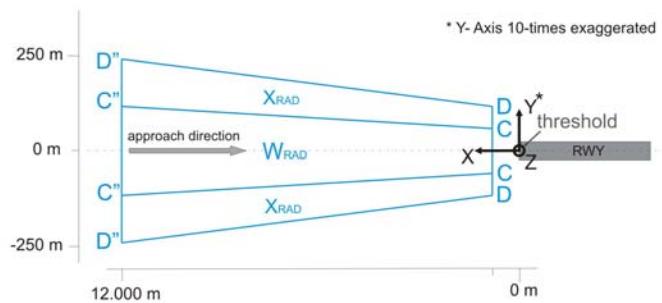


Figure 9. determined ANP-based OAS final approach surface – top view

The blue lines in Fig. 9 are marking the edges of OAS surfaces as determined. The corners on each side were named according to PANS-OPS declaration with the capital letters C and D for the threshold nearest corners and with C'' and D'' for the opposite corners. For clarification purposes, Tab. IV below summarizes the corner coordinates, were the origin of the coordinate system is located on the threshold (see also Fig. 9) with z=0 at threshold elevation:

TABLE IV. COORDINATES OF DETERMINED OAS SURFACES USING TLS OF 1×10^{-8} PER APPROACH

Corner point	Corner Point Coordinates		
	X	Y	Z
C	543 m	+/- 52.93 m	0 m
D	543 m	+/- 105.94 m	43.69 m
C''	12'000 m	+/- 122.03 m	533.12 m
D''	12'000 m	+/- 243.18 m	644.13 m

As seen the determined OAS surfaces are not ending at the threshold (line D-C-C-D in Fig. 9). The threshold distance at the end of surfaces is 543 m, this is the point were the W-surface is at the same altitude as the landing threshold and therefore the ground collision risk reached the TLS. This can be seen as the obstacle clearance height (OCH) for an obstacle free environment (ground collision risk determines the OCH), which is about 44 m (143 ft) in this case.

E. Impacts of using lower TLS

When assuming a TLS of 1×10^{-8} per approach instead of 1×10^{-7} per approach the shown OAS surfaces will increase slightly. This increase is surely depending on the threshold distance due to angular characteristic. For near threshold distances (less than 1000 m), this increase will be about 3 m in vertical direction and about 6 m in lateral direction. For far threshold distances (12'000 m), these values will increase about 8 m respectively 16 m (for comparison, see Table IV). Consequently, the OCH will increase from about 44 m to about 47 m. Altogether, these increases are very low, as the percentage of surface volume increase is only about 13.4 % when using this 10-times lower TLS.

VI. CONCLUSION

Design of specific Obstacle Assessment Surfaces for current and upcoming approach procedures is crucial for future airport development when benefits of improved navigational performance shall be fully exploited. This paper presented a method to determine navigational performance during the final approach phase and a strategy for calculating ANP-based OAS executed for ILS final approach. This method may easily be transferred to innovative RNP/RNAV approach procedures, when respective reference data becomes available.

Nevertheless, the results of a statistical approach path deviation analysis for RNAV procedures may be different to the here shown results for ILS approach (see Fig. 9). Due to angular signal characteristics of ILS radio signals, an increasing approach funnel was found. Based on the non-angular characteristic of RNP/RNAV procedures this funnel is assumed to convert into a tube covering the entire final approach segment. Consequently, geometric advantages of this design concept appear with increasing threshold distance.

Finally, assuming a similar distribution behavior for RNP/RNAV approaches as for ILS approaches, a lower TLS of 1×10^{-8} per approach could be a proper way to reach SESAR goal of increasing in safety by a factor of 10, as increasing size of OAS surfaces is very low.

REFERENCES

- [1] SESAR Consortium, "European Air Traffic Management Master Plan", Edition 1, 30.03.2009
- [2] Boeing Commercial Airplane Group, "Statistical summary of commercial jet airplane accident, worldwide operations 1959 – 2008", Seattle, July 2009
- [3] ICAO, "Procedures for Air Navigation Services – Aircraft Operations (PANS-OPS), Volume I, Doc. 8168", 5th Edition, Montreal, 2006
- [4] H. Fricke, "Integrated collision risk modelling - CoRiM: An airborne safety and protection system for upcoming ATM concepts", ISBN 3-89574-418-2, Berlin, 2001
- [5] M. Vogel, C. Thiel and H. Fricke: "A quantitative Safety Assessment Tool based on Aircraft Actual Navigation Performance", International Conference on Research in Airport Transportation (ICRAT), Budapest, 2010
- [6] ICAO, "Performance-based Navigation (PBN) Manual, DOC 9613-AN/937", 3rd Edition, Montreal, 2008
- [7] Eurocontrol, "Guidance Material for the Design of Terminal Procedures for Area Navigation (DME/DME, B-GNSS, Baro- VNAV & RNP-RNAV)", 3rd Edition, March 2003
- [8] C. Thiel, „Untersuchung der Navigationsgenauigkeiten von Luftfahrzeugen im Flughafennahrbereich - Modellierung und Implementierung dreidimensionaler Verteilungsfunktionen unter Berücksichtigung der Navigationsinfrastruktur an großen Verkehrsflughäfen“, Diploma Thesis, TU Berlin, Berlin, June 2009
- [9] M. Kietzmann and H. Fricke, „Entwicklung eines quantitativen Bewertungsmodells zur objektiven Beurteilung des Einflusses neuer Betriebsverfahren auf die Sicherheit im Luftverkehr,“ DGLR Jahrestagung 2004, Dresden, Germany
- [10] ICAO, "Annex 10 - Aeronautical Telecommunications, Volume 1", 6th Edition, Montreal, 2006
- [11] ICAO, "Manual on the Use of the Collision Risk Model (CRM) for ILS Operations, DOC 9274- AN/904", 1st Edition, Montreal, 1980
- [12] ICAO, „Airport Services Manual, Part 6 – Control of Obstacles, DOC 9137", 2nd Edition, Montreal, 1983
- [13] ICAO, "Annex 14 – Aerodromes, Volume I, Aerodromes Design and Operations", 5th Edition, Montreal, July 2009
- [14] H. Schulz: "Entwicklung von Verfahrensschutzraum für satellitengestützte Präzisionsanflüge mit bodenseitiger Augmentierung", Dissertation, TU Berlin, Berlin, April 2004

Comparison of Arrival Tracks at Different Airports

Yimin Zhang, Ph.D. Student

Systems Engineering and Operations Research
Center for Air Transportation Systems Research
Fairfax, VA
yzhangk@gmu.edu

John Shortle, Lance Sherry

Systems Engineering and Operations Research
Center for Air Transportation Systems Research
Fairfax, VA
jshortle@gmu.edu, lsherry@gmu.edu

Abstract—This paper analyzes arrival flight track data at Chicago (ORD) and Atlanta (ATL) airports. We investigate distributions of vertical and lateral position at different distances from the runway threshold in instrument meteorological conditions (IMC) and visual meteorological conditions (VMC). In IMC, the observed standard deviations at different distances are similar between the two airports. The results reported in this paper are also similar to those reported at St. Louis (STL) in [1]. Visual comparison of the observed distributions also shows a close similarity. This provides some indication that distributions observed at one airport in IMC may generalize to other airports. In VMC, there are some differences between the distributions. We also fit probability density functions (PDFs) to lateral and vertical positions. In general, the normal distribution provides the best fit among the normal, lognormal, gamma, and Weibull families. The quality of the fit is better in IMC closer to the runway threshold.

Keywords- flight tracks; probability density functions

I. INTRODUCTION

The objective of this paper is to measure properties of flight tracks at two major U.S. airports and compare the results. Since it is challenging to measure flight tracks at all airports, a natural question arises – are the flight tracks at one airport representative of the flight tracks at other airports? If similar statistical properties are observed at different airports, this provides evidence that statistical properties observed at one airport may possibly be “extrapolated” to other airports. Such evidence does not prove such an assertion, but simply lends evidence in that direction. On the other hand, if statistical properties at different airports are different, this provides evidence that airports may need to be individually measured.

This paper compares arrival flight track data at two major U.S. airports, Chicago (ORD) and Atlanta (ATL). The basic observation is that observed standard deviations are reasonably similar between the two airports during instrument meteorological conditions (IMC). The results in this paper are also similar to results presented in [1] for STL. (The results in [1] were obtained from an ASDE-X system that was augmented by off-airport multilateration sensors to enhance the accuracy of measurements along the arrival corridors.) The similarity of the results provides some preliminary indication that statistical properties of flight tracks may be similar at other major U.S. airports during IMC.

II. METHODOLOGY

This paper uses an existing algorithm and methodology for processing track data. Initial and updated versions of the algorithms are given in [2] and [3]. These previous papers have analyzed multilateration data at DTW. Here, we apply the algorithms to flight track data for ORD and ATL. We have been able to apply the existing algorithms to this analysis with few modifications.

Another study that has analyzed distributions associated with flight tracks on approach is [4]. Other studies that have analyzed statistical distributions of aircraft separations on arrival, both in terms of distance and time are [5-12]. Statistical measurements of position deviations have also been made in the en-route environment (e.g., [13-15]).

III. DATA SUMMARY

This paper uses pre-processed flight track data at ATL and ORD as a basis for analysis of flight tracks. Flight track data is typically generated from an ASDE-X system that synthesizes data from a number of sources including multilateration sensors and surveillance radar. The intended coverage area includes the airport surface as well as arrival corridors out to 5 nm from runway thresholds. Accuracy on the surface is 20 feet (one standard deviation) or better. Away from the threshold within the coverage area, accuracy is as good as or better than existing sensors (e.g., surveillance radar).

The data for each day consist of several comma-separated text files, each containing a single table with the following fields: aircraft ID, time, x-coordinate, y-coordinate, altitude, aircraft type (manufacturer and model), and wake category. The wake category (heavy, B757, large, small) is inserted using a mapping table created from previous work based on the aircraft type. The origin of the coordinate system is the airport control tower (Fig. 1 shows the ORD airport diagram). The x/y-coordinate measurements are based in part on multilateration data. The altitude measurements are based on barometric pressure and not on multilateration measurements. The fields must be modified slightly in order to use the algorithms developed in previous studies [2,3].

We analyze arrivals at ATL, runways 9R, 27L, 8L, and 26R, and arrivals at ORD, runways 10, 28, 4R, and 14R. These are main runways for arrivals based on the most commonly used configurations. In total, 39,278 arrivals are

observed (22 days for ATL and 31 days for ORD). Some arrivals are removed due to various data-integrity checks (track is too short, gaps in the track, etc.). This leaves 31,566 arrivals viewed as valid to be analyzed. Table 1 shows a summary of flight tracks observed for all runways under both IMC and VMC.

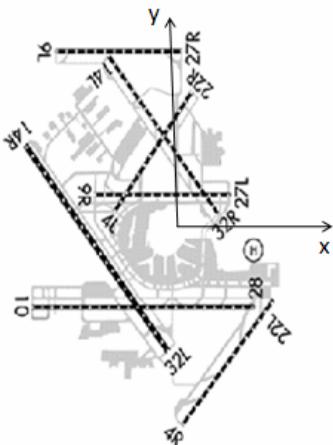


Figure 1. ORD airport diagram (www.airnav.com)

TABLE I. TRACKS SUMMARY(COUNTS)

Airport	Runway	IMC	VMC	TOTAL
ATL	8L	1167	3837	5004
	9R	1169	3287	4456
	26R	469	3422	3891
	27L	589	3935	4524
ORD	10	527	4629	5156
	28	198	2335	2533
	4R	180	3132	3312
	14R	207	2483	2690

For illustration purposes, Table 2 shows an example of the processing steps and data for one day of data at ATL, runway 9R. (See [2,3] for more details on these steps.) The first row is the number of data points (rows) in the original file. The second row is the number of points remaining after discarding points outside of a defined box. Points outside of the box are assumed to belong to operations on other runways. The coordinates of the box are specific to the runway being investigated. The third row is the number of distinct arrival tracks extracted from the data. The final row is the number of tracks remaining after discarding tracks that fail a data quality check (e.g., the tracks are too short or there are gaps in the data).

TABLE II. DATA SUMMARY (ONE DAY, ATL 9R)

# of points in original file	4,730,919
# of points after boxing	139,959
# of candidate tracks	478
# of valid tracks	473

IV. ANALYSIS OF FLIGHT TRACKS

The weather conditions (VMC and IMC) associated with a track are defined by comparing the time of the first point of the track with airport weather information in the Aviation System Performance Metrics (ASPM) database. Fig. 2 shows a top-level view of all 527 flight tracks at ORD, runway 10 in IMC. In IMC, aircraft fly through the final approach fix (approximately 5 nm from the threshold) straight to the runway. As expected, the lateral positions converge to the centerline of the runway as aircraft get closer to the threshold.

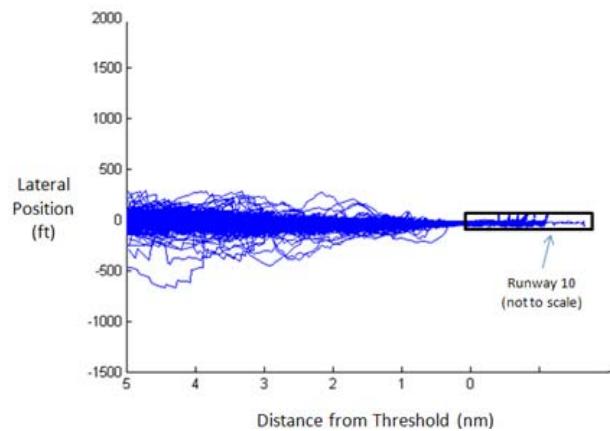


Figure 2. Top-level view of all tracks, ORD, runway 10, IMC

Fig. 3 compares the lateral distribution in VMC and IMC at 4 nm from the threshold of runway 10 at ORD. The IMC distribution is concentrated more closely about the centerline, while the VMC distribution has a larger number of points in the tails of the distribution. In VMC, it is possible for aircraft to curve in after the final approach fix, which leads to heavier tails of the VMC distribution. Similar results are seen at other distances from the threshold, and at ATL.

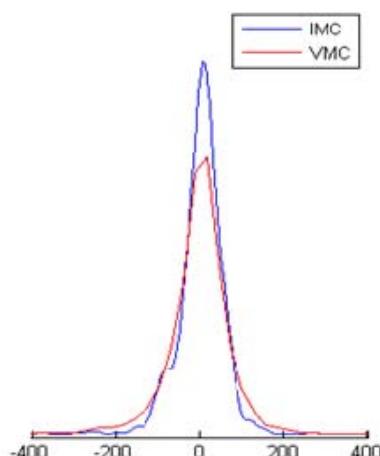


Figure 3. Density of lateral position

To obtain distributional fits for the data, we first break the data down by airport (ATL/ORD), by weather conditions (VMC/IMC), and by distance to the threshold (0, 1, 2, 3, 4, 5 nm). In each case, we determine the best fit among the following families of distributions: normal, lognormal, Weibull, and gamma. The best fit is defined with respect to minimizing the maximum likelihood estimation. Fig. 4 shows an example of the probability density function (PDF) fits at ATL, 3 nm from the threshold, during IMC. (The data for all runways at ATL are combined in these results.) The best fit in this case is the normal distribution. The gamma distribution is the second best fit.

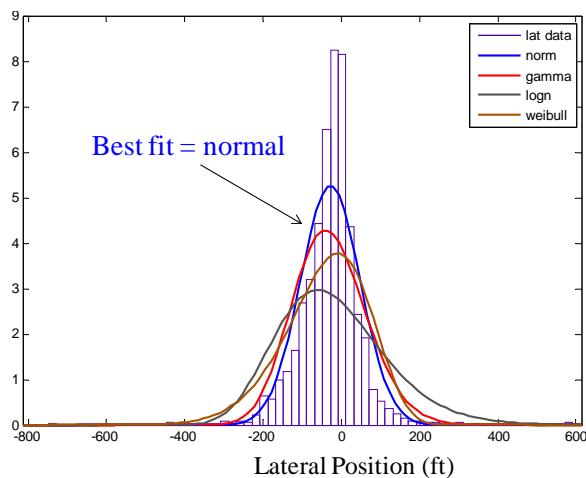


Figure 4. Sample PDF fits for lateral position, ATL, IMC

Fig. 5 shows an example comparing the distributional fits at several distances from the threshold. In each case, the normal distribution (the blue curve) provides the best fit among the four candidate families. Note that the scale is not exactly the same among the three graphs. The purpose is to illustrate the PDF fits relative to the observed histograms, rather than to provide the absolute values. Close to the runway threshold, the normal distribution provides a better fit. At further distances from the threshold, the quality of the fit is not

as good, though the normal distribution is still the best fit among the four candidate distributions. In particular, at further distances, the observed distribution is more narrow in the body than what would be predicted by a normal distribution. The observed extreme values are also more extreme than what would be predicted by a normal, though this is easier to see in the next figure. Similar results are observed for ORD.

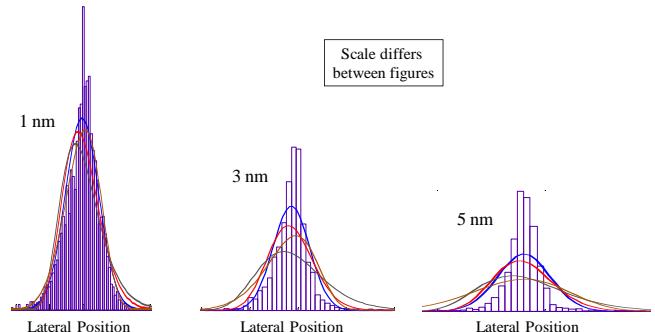


Figure 5. Sample PDF fits for lateral position, ATL, IMC

Fig. 6 shows a plot of the observed data versus quantiles of a standard normal distribution (a quantile-quantile plot). If the data were normally distributed, the data would fall on a straight line. In the figure, the data are nearly normally distributed, since most of the data are close to the straight line. The end points are above the line on the right and below the line on the left. This means that the extreme points of the observations are larger in magnitude than what would be predicted by a normal distribution. That is, the tails are slightly heavier than a normal distribution.

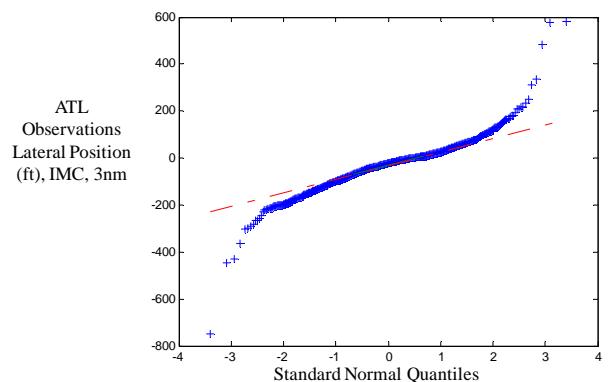


Figure 6. Observed lateral position versus normal quantiles, ATL, IMC, 3nm

While we do not show all combinations of results here, the normal distribution is typically the best fit among the four candidate distributions. This is particularly true in IMC and closer to the runway threshold. The quality of the fit typically degrades at further distances from the threshold. The tails of the observed distributions are typically heavier than what is predicted by the normal fit. That is, similar to Fig. 6, analogous plots at different distances at both airports show similar trends in the tails of the distributions.

V. COMPARISON BETWEEN AIRPORTS

In this section, we compare results between the two airports ATL and ORD. We do not report any results using DTW data from previous studies, because the accuracy of that data (collected in 2003) may not be sufficient for analysis significantly beyond the runway threshold.

As an initial comparison, Table 3 shows the standard deviation of lateral position of aircraft in IMC. Our results are similar to those reported at St. Louis (STL) in [1] as well. Note that there are some differences in the data sources between this paper and [1]. In [1] (STL), the ASDE-X system is augmented by off-airport multilateration sensors to enhance the accuracy of measurements along the approach corridor. In our data (ATL and ORD), there is no such augmentation, though the intended coverage area includes arrival corridors out to 5 nm from the threshold with an accuracy equal to or better than existing radar. Thus, in principle, the data accuracy in this paper may not be as good as in [1]. However, the similarity of values in Table III indicates that the precision may be good enough for the purpose of evaluating the variability in the tracks.

A second difference is that weather conditions in this paper are defined based on the VMC/IMC flag in the ASPM database. In [1], the weather conditions are defined using Automated Surface Observing System (ASOS) data, where “hard IMC” at STL means less than 4 statute miles in visibility or less than 1,200 ft ceiling.

TABLE III. STANDARD DEVIATION OF LATERAL POSITION (FT)

	Runway 8L at ATL, IMC	Runway 10 at ORD, IMC	STL, from [1], “hard IMC”
Distance from threshold (nm)	Standard deviation (ft)	Standard deviation (ft)	Standard deviation (ft)
0	8	7	14
1	20	22	20
2	35	34	30
3	60	47	44
4	61	63	50
5	110	58	66

The observed standard deviations reported in Table III are similar across each row, except possibly in the 5-nm case. We investigate this case further. Fig. 7 compares the observed distributions of lateral position at 5 nm in IMC. Visually, the distributions look nearly identical. One difference is that ATL has several larger observations that extend beyond 400 feet from the centerline, whereas ORD has fewer such observations (though this is difficult to see from the figure). These large observations greatly increase the observed standard deviation, which accounts for the differences in the standard deviation between ATL and ORD and STL. This illustrates why moment-based measures, which are sensitive to large observations, can be misleading. Aside from these points, the distributions are visually nearly identical in the figure.

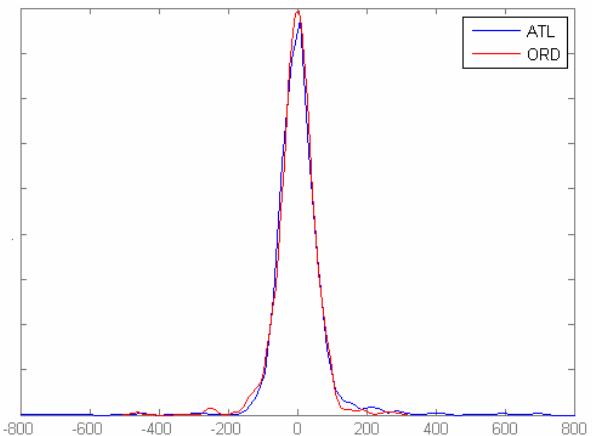


Figure 7. Comparison of lateral distribution, 5 nm, IMC

Table 4 gives the standard deviation of vertical position of aircraft landing on both runways under IMC. We do not include data at the threshold since altitude is only reported in increments of 25 feet, so the lack of precision dominates the estimate of standard deviation at the threshold. In [1], the standard deviations are reported only at a subset of the distances. Again, there is general agreement in the reported values.

TABLE IV. STANDARD DEVIATION OF VERTICAL POSITION (FT)

	Runway 8L at ATL, IMC	Runway 10 at ORD, IMC	STL, from [1], “hard IMC”
Distance from threshold (nm)	Standard deviation (ft)	Standard deviation (ft)	Standard deviation (ft)
1	26	26	30
2	33	31	
3	48	40	50
4	55	61	
5	64	76	118

Figs. 8 and 9 show a comparison in VMC of the observed standard deviations for lateral and vertical positions. The figures include data from ATL, runway 8L, and ORD, runway 10 (no results are reported in [1] for VMC). For vertical position, the results are nearly identical. For lateral position, there is a difference between the standard deviations at the two airports / runways. This might be because pilots have more flexibility in visual conditions regarding the path flown to the runway, so the path may depend on a number of airport/runway-specific factors such as the geometry of the location to enter the terminal airspace.

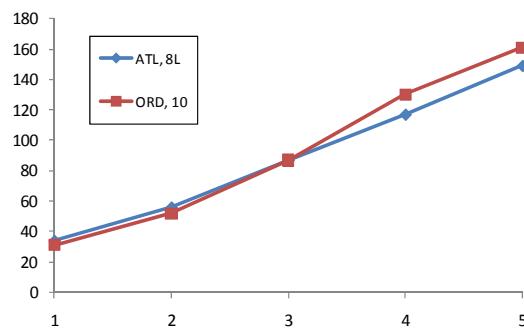


Figure 8. Standard deviation of vertical position in VMC

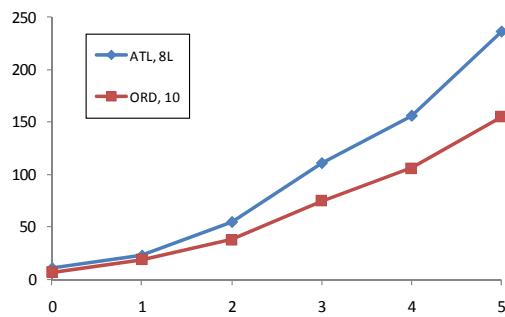


Figure 9. Standard deviation of lateral position in VMC

CONCLUSIONS

This paper analyzed arrival flight tracks from ASDE-X data at Chicago (ORD) and Atlanta (ATL) airports. We investigated distributions of vertical and lateral position at different distances from the runway threshold in instrument meteorological conditions (IMC) and visual meteorological conditions (VMC). In IMC, the observed standard deviations were similar between the two airports. The results were also similar to those reported at St. Louis (STL) in [1], based on data from an ASDE-X system augmented with off-airport sensors to increase accuracy of measurements in the arrival corridors. Visual comparison of the observed distributions at ATL and ORD also showed a close similarity. This provides some indication that distributions observed at one airport in IMC may generalize to other airports. This does not prove such an assertion, but simply provides some evidence in that direction. In VMC, some differences were observed in lateral position between different runways at different airports. This might be expected, since there is more flexibility in the flight path in visual conditions. Thus, it may be more difficult to generalize VMC distributions from one airport to others.

We also fit probability density functions (PDFs) to lateral and vertical positions. In general, the normal distribution provided the best fit among the normal, lognormal, gamma, and Weibull families. The quality of the fit was better in IMC

closer to the runway threshold and not as good at further distances from the threshold.

ACKNOWLEDGMENT AND DISCLAIMER

This work has been supported by NASA and Northwest Research Associates (NWRA) through sub-agreement #NWRA-08-S-114 and Integrated Airport Capacity Management (IACM) project with Metron Aviation. The opinions and conclusions in this paper are solely those of the authors and not necessarily those of NWRA or NASA. The authors thank Rafal Kicinger (Metron Aviation, Inc.) and Vivek Kumar (CATSR, GMU) for help in analyzing the data.

REFERENCES

- [1] Hall, T., M. Soares. 2008. Analysis of localizer and glide slope flight technical error. 27th Digital Avionics Systems Conference, St. Paul, MN.
- [2] Juddi, B., J. Shortle, and L. Sherry. Statistics of the approach process at Detroit Metropolitan Wayne County Airport. In *Proceedings of the International Conference on Research in Air Transportation*, Belgrade, Serbia and Montenegro, 2006, pp. 85-92.
- [3] Shortle, J., Y. Zhang, J. Wang. 2010. Statistical characteristics of arrival flight tracks. In *Proceedings of the Transportation Research Board Annual Meeting*, Washington, DC.
- [4] Levy, B., P. Som, B. Greenhaw. 2003. Analysis of flight technical error on straight, final approach segments.
- [5] Levy, B., J. Legge, and M. Romano, Opportunities for improvements in simple models for estimating runway capacity, 23rd Digital Avionics Systems Conference, Salt Lake City, UT, 2004.
- [6] Boswell, S. 1993. Evaluation of the capacity and delay benefits of terminal air traffic control automation. DOT/FAA/RD-92/28, MIT Lincoln Laboratory.
- [7] Lebron, J. 1987. Estimates of potential increases in airport capacity through ATC system improvements in the airport and terminal areas. FAA-DL5-87-1.
- [8] Haynie, C. 2002. An investigation of capacity and safety in near-terminal airspace guiding information technology adoption. Ph.D. Dissertation, George Mason University.
- [9] Xie, Y. 2005. Quantitative analysis of airport arrival capacity and arrival safety using stochastic models. Ph.D. Dissertation, George Mason University.
- [10] Ballin, M., and H. Erzberger. An Analysis of Landing Rates and Separations at the Dallas / Fort Worth International Airport. NASA Technical Memorandum 110397, 1996.
- [11] Andrews, J., J. Robinson. 2001. Radar-based analysis of the efficiency of runway use. AIAA Guidance, Navigation & Control Conference, Montreal, Quebec. AIAA-2001-4359.
- [12] Rakas, J., and H. Yin. Statistical Modeling and Analysis of Landing Time Intervals: Case Study of Los Angeles International Airport, California. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1915, TRB, National Research Council, Washington, D.C., 2005, pp. 69-78.
- [13] Harrison, D. 1987. Some preliminary results of estimating the probability of vertical overlap from the distribution of single aircraft deviations from North Atlantic Traffic. UK CAA report.
- [14] Campos, L., J. Marques. 2002. On safety metrics related to aircraft separation. *Journal of the Royal Naval Society*, 55:39-63.
- [15] Campos, L., J. Marques. 2004. On a combination of gamma and generalized error distributions with applications to flight path deviations. *Communications in Statistics: Theory and Methods*, 33(10): 2307-2332.

Stochastic Validation of ATM Procedures by Abstraction Algorithms

Maria D. Di Benedetto, G. Di Matteo and A. D’Innocenzo

Department of Electrical and Information Engineering, Center of Excellence DEWS
University of L’Aquila, Italy

e-mail: mariadomenica.dibenedetto@univaq.it, giuliadimatteo@hotmail.it, alessandro.dinnocenzo@ing.univaq.it

Abstract—In this paper we propose a methodology for formal reasoning based on stochastic hybrid systems theory and abstraction algorithms for stochastic dynamical systems, which provides a powerful framework to analyze stochastic models of ATM procedures. We propose the use of automatic tools for verifying probabilistic properties of ATM scenarios. In particular, we propose to use PCTL logic to define probabilistic properties of interest. We address a simple single-agent procedure of the *A³*(Autonomous Aircraft Advanced) *ConOps* (Concept of Operations), describe a dynamical model for the aircraft deterministic dynamics and for the wind stochastic dynamics, and used MATLAB and PRISM tools in order to perform stochastic analysis of properties of interest of the addressed scenario.

Index Terms—Air traffic management, Stochastic hybrid systems, Abstraction algorithms, Probabilistic model checking.

I. INTRODUCTION

The introduction of new Air Traffic Management (ATM) procedures is a necessary condition for achieving safety and efficiency objectives requested by the increasing air traffic. Modeling, simulation and formal analysis and validation of new ATM procedures is an important and necessary step for the development of ATM systems. In the context of the iFly project, our research focuses on development of novel concepts and technologies for addressing the issues discussed above, in order to provide automatic tools for the ATM systems under development and standardization.

In the past, the Air Traffic Controller (ATC) and the pilots had access to different data, and the responsibility of changes in procedures and operations were totally delegated to the ATC. The introduction of the next generation of ATM systems forecasts the use of ground and on-board integrated surveillance systems, which guarantee a cooperation between the ATC and the pilots. Moreover, new technologies provide broadcast communication of an aircraft position in the airspace, thus enabling the possibility of decentralization of decision making from the ATC to the pilot. These are the enabling technologies for development of a plethora of applications, such as the Airborne Separation Assistance System (ASAS) [1], which aims to improve efficiency of air traffic management procedures by a decentralization of responsibility among the ATC and the pilots.

The more advanced ASAS application is the Airborne Self Separation (ASEP) [2], which aims to a total shift of responsibility to the pilots flying in a specified airspace. Within this airspace, the pilots are responsible of maintaining

safety separation with the other aircraft using the on board surveillance system. The operative concept ConOps [3] consists of two planning phases and one validation phase. The first planning phase derives from the Autonomous Aircraft Advanced (*A³*) concept [3], which contemplates a network of aircraft, each responsible of Airborne Self Separation with no ATC ground support. The second phase contemplates analysis and validation of results of the first phase, in order to improve the *A³* concept, including the ATC support when necessary. These new concepts are a potential solution to the increasing air traffic density expected in the future years, and forecast an increase of safe air traffic from three to six times the current air traffic. The main problem is providing guarantee that the new air traffic procedures are sufficiently safe.

In this paper we propose to apply a methodology for formal reasoning based on stochastic hybrid systems theory, that provides a powerful framework to analyze multi-agents stochastic models of ATM procedures. We propose the use of automatic tools for verifying probabilistic properties of ATM scenarios. In particular, we propose to use PCTL logic to define probabilistic properties of interest (we refer to [4] and references therein for a survey on PCTL). Recently, formal verification of stochastic models has been transformed from an academically attractive discipline to a research effort prone to yield industrially relevant applications, and tools for probabilistic model checking have been developed: we propose the use of PRISM [5], [6], [7] for automatic verification of PCTL properties on ATM procedures.

However, the dynamical analysis of high-dimensional, stochastic models poses a number of challenges. When direct analysis of the model under study is impaired by its sheer complexity, automatic verification and algorithmic control design procedures are essential. An approach that is successfully used to cope with the issue of computational complexity and scalability is that of *abstraction*: a system with smaller state space is sought, which is *equivalent* to the original system. System equivalence implies that some properties of the original (complex, possibly infinite dimensional) system are preserved by the (simple, possibly finite dimensional) abstraction. For this reason, the property of interest can be efficiently checked on the abstraction, in finite time and/or with a lower computational complexity. Figure 1 illustrates the main phases of our verification algorithm we propose.

In the first block, a detailed continuous-time stochastic model of the ATM procedure (e.g. a stochastic model of

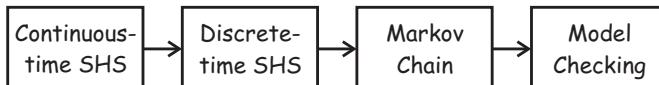


Fig. 1: Verification algorithm flow.

the aircraft dynamics) is defined, using the mathematical framework of continuous time Stochastic Hybrid Systems (ct-SHS). This model can be discretized with respect to the time variable, thus obtaining a discrete time SHS (dt-SHS). We refer to [8] and references therein for the formal definition of discrete and continuous time SHSs. In the third block, a Markov Chain abstraction of the model is obtained using the abstraction procedure proposed in [9]. This abstraction procedure is essentially a partition of the state space, which depends on a tunable parameter δ (the width of the partition grid). The reason for using this abstraction is that it provides an approximation of the original system, and it can be used to perform automatic model checking using the tool PRISM. The results of the model checking verification directly apply to the original system, modulo an approximation error ϵ . This approximation error ϵ can be chosen a-priori, by modifying the parameter $\delta(\epsilon)$ of the abstraction procedure from dt-SHS to Markov Chain.

The paper is organized as follows. In Section II we describe the ATM scenario we considered to apply our methodology, i.e. the Hole in the clouds scenario illustrated in [3]. In Section III we describe a dynamical model for the aircraft deterministic dynamics, and for the wind stochastic dynamics. In Section IV we present our simulation results.

II. SCENARIO

We illustrate a simple procedure of the A^3 (Autonomous Aircraft Advanced) *ConOps* (Concept of Operations), i.e. the Hole in the clouds scenario illustrated in [3].

In A^3 ConOps the concept of airspace has been redefined, introducing the concept of Performance Based Airspace (PBA). A^3 airspace is divided into 3 categories, as illustrated in Figure 2: the *Managed Airspace* (MA) is a high-density area; the *Unmanaged Airspace* (UA) is an area where ATC services are not accessible; the *Performance Based Airspace* (PBA) is an airspace whose boundaries are defined in time and space through MA and UA dynamic assignment.

In PBA autonomous aircraft are responsible for separation, according to the AFR (Autonomous Flight Rules). Operations are usually conducted under AFR or IFR (Instrument Flight Rules), while operations under VFR (Visual Flight Rules) are only admitted at specific altitudes. In PBA airspace aircraft have to guarantee self-separation and safe manoeuvres. Any conflict has to be avoided using appropriate manoeuvres: the final objective is safe cruise avoiding any conflict, e.g. *Protected Airspace Zones* (PAZ), *Restricted airspace areas* (RAA), or *Weather hazards areas* (WHA).

We consider an A^3 flight, defined as the flight between a departing Terminal Control Area (TMA) exit point, and an arriving TMA entry point, constrained by a Controlled Time

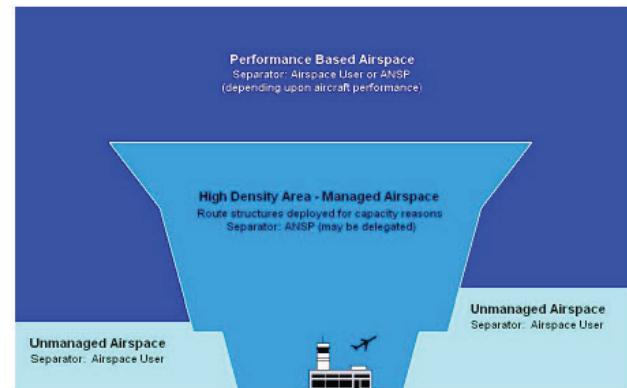


Fig. 2: Airspace classification, from [3].

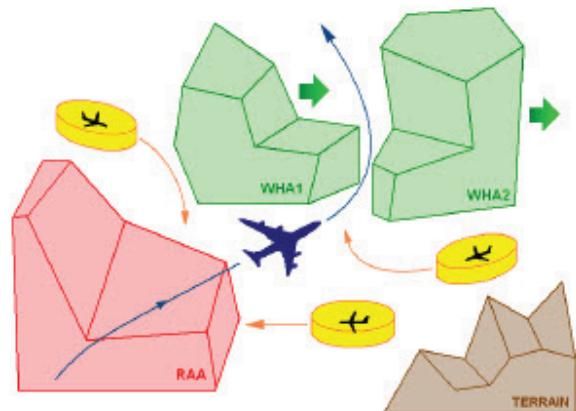
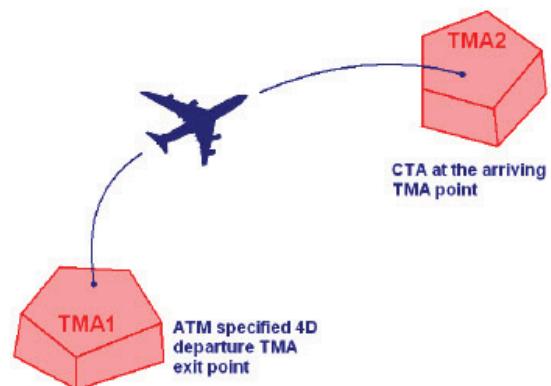


Fig. 3: Conflict environment in PBA, from [3].

of Arrival (CTA) at the arriving TMA entry point, as illustrated in Figure 4.

Fig. 4: A^3 flight, from [3].

During the flight, the aircraft follows its Business Trajectory (RBT) and maintains separation from other aircraft and conflicts, respecting constraints imposed by Traffic Flow Management. Given a Business Trajectory, we apply our methodologies to verify position and time of arrival to the arriving TMA without entering conflict areas, by taking into

account stochastic wind disturbance on the aircraft dynamics. Figure 5 illustrates a scenario, where WHA conflicts are present.

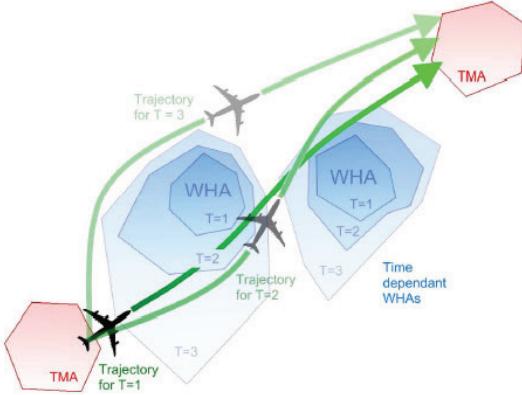


Fig. 5: WHA conflicts A^3 flight, from [3].

III. AIRCRAFT DYNAMICAL MODEL

We use a Point Mass Model (PMM) for modeling aircraft dynamics. We denote by X, Y horizontal position, by h altitude, by V true airspeed, by γ flight trajectory angle, and by ψ heading angle. Wind is considered as a disturbance on the aircraft dynamics, and is modeled by its speed $W = (w_1, w_2, w_3) \in \mathbb{R}^3$.

We use the following dynamical model from [10]:

$$\begin{aligned}\dot{X} &= V \cos(\psi) \cos(\gamma) + w_1 \\ \dot{Y} &= V \sin(\psi) \cos(\gamma) + w_2 \\ \dot{h} &= V \sin(\gamma) + w_3 \\ \dot{V} &= \frac{1}{m} [(T \cos(\alpha) - D) - mg \sin(\gamma)] \\ \dot{\psi} &= \frac{1}{mV} (L \sin(\phi) + T \sin(\alpha) \sin(\phi)) \\ \dot{\gamma} &= \frac{1}{mV} [(L + T \sin(\alpha)) \sin(\phi) - mg \cos(\gamma)]\end{aligned}\quad (1)$$

where T denotes engine thrust, α attack angle, ϕ yaw angle, m aircraft mass and g gravity acceleration. L and D respectively denote lift and drag forces, which are functions of the state and the attack angle. Typically:

$$\begin{aligned}L &= \frac{C_L S \rho}{2} (1 + c\alpha) V^2, \\ D &= \frac{C_D S \rho}{2} (1 + b_1 \alpha + b_2 \alpha^2) V^2,\end{aligned}$$

where S denotes wing surface, ρ air density, and C_D, C_L, c, b_1, b_2 lift and drag aerodynamic coefficients that depend on the flight phase.

Figure 6 illustrates how forces act on the aircraft in the model described above.

From 1 we derive a 6-dimensional model of the aircraft $x = (x_1, x_2, x_3, x_4, x_5, x_6)^T \in \mathbb{R}^6$, with 3 constant inputs $u = (u_1, u_2, u_3)^T \in \mathbb{R}^3$ and 3 disturbance components $w = (w_1, w_2, w_3)^T \in \mathbb{R}^3$. By defining $x_1 = X$, $x_2 = Y$, $x_3 = h$, $x_4 = V$, $x_5 = \psi$, $x_6 = \gamma$, $u_1 = T$, $u_2 = \phi$, $u_3 = \gamma$, and considering the consumption coefficient η , we obtain the following dynamics:

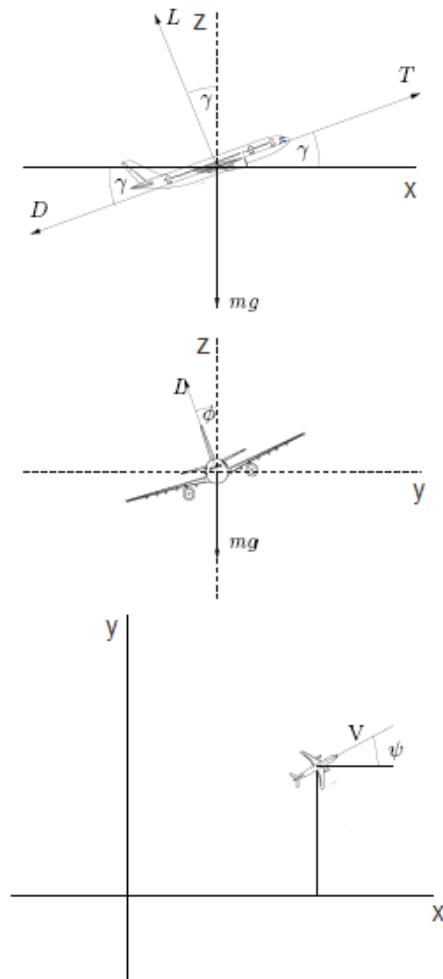


Fig. 6: Forces acting on the aircraft.

$$\dot{x} = \begin{bmatrix} x_4 \cos(x_5) \cos(u_3) + w_1 \\ x_4 \sin(x_5) \cos(u_3) + w_2 \\ x_4 \sin(u_3) + w_3 \\ -\frac{C_D S \rho}{2} \frac{x_4^2}{x_6} - g \sin(u_3) + \frac{1}{x_6} u_1 \\ \frac{C_L S \rho}{2} \frac{x_4}{x_6} \sin(u_2) \\ -\eta u_1 \end{bmatrix} \quad (2)$$

State and input are subject to the following constraints: $x_3 > 0$, $x_4 \in [V_{min}, V_{max}]$, $x_6 \in [m_{min}, m_{max}]$, $u_1 \in [T_{min}, T_{max}]$, $u_2 \in [\phi_{min}, \phi_{max}]$, $u_3 \in [\gamma_{min}, \gamma_{max}]$. Values for state and input constraints and for parameters C_D , S and ρ can be found from the database BADA (Base of Aircraft Data) [11].

Wind is modeled by a nominal component and a stochastic component $w = w_n + w_s$. The stochastic component is modeled by Gaussian random variables, i.e. by a random field $w_s : \mathbb{R} \times \mathbb{R}^3 \rightarrow \mathbb{R}^3$, where $w_s(t, P)$ represents wind in point $P \in \mathbb{R}^3$ at time $t \in \mathbb{R}$. We assume that $w_s(\cdot, \cdot)$ satisfies the following properties:

- 1) $w_s(t, P)$ is a Gaussian random variable with mean

$\mu(t, P)$ and covariance matrix $\Sigma(t, P)$.

- 2) The random field is isotropic in x e y , i.e. the correlation structure does not change for rotations in the horizontal plane.
- 3) $w_{s1}(t, P), w_{s2}(t, P), w_{s3}(t, P)$ are independent $\forall t \in \mathbb{R}, \forall P \in \mathbb{R}^3$.

In the addressed scenario, we assume that altitude h , true airspeed V and heading angle ψ are constant, and that γ is equal to zero. Under these assumptions, equations (2) assume the following form:

$$\dot{x} = \begin{bmatrix} V\cos(\psi) + w_n\cos(\beta) + w_{s1} \\ V\sin(\psi) + w_n\sin(\beta) + w_{s2} \end{bmatrix} \quad (3)$$

where $x = (x_1, x_2)^T \in \mathbb{R}^2$ model the aircraft position in the plane $x_1 = X, x_2 = Y$, $(w_n\cos(\beta), w_n\sin(\beta))^T \in \mathbb{R}^2$ is the deterministic component of the wind where w_n e β are respectively wind velocity and direction, and $w_s = (w_{s1}, w_{s2})^T \in \mathbb{R}^2$ is the stochastic component of the wind. Using these dynamics, it is possible to derive a continuous time SHS describing the dynamics of the aircraft. Starting from this model and choosing a sampling time Δ , we derive a discrete time SHS by applying the classical Eulero-Maruyama discretization with constant step Δ .

From equations in (3) we obtain the deterministic component f that characterizes the dynamics of the aircraft for the dt-SHS model:

$$f = \begin{pmatrix} V\cos(\psi) + w_n\cos(\beta) \\ V\sin(\psi) + w_n\sin(\beta) \end{pmatrix}$$

We assume the aircraft is flying at cruise speed and at flight level 350. The table 7 (obtained from [11]) reports aircraft data in cruise, climb and descent phase.

At flight level 350, in cruise phase, the corresponding true airspeed (TAS) is 461 kts (853.772 km/h). We assume that $\psi = \pi/4$, and that the wind has a constant deterministic component with speed 50 km/h and direction from North to East.

We resume the parameters used in our simulations: $V = 853.772 \text{ km/h} = 0.2372 \text{ km/s}$, $\psi = \pi/4$, $w_n = 50 \text{ km/h} = 0.0139 \text{ km/s}$, $\beta = 3\pi/8$.

IV. SIMULATION RESULTS

Given the aircraft and wind dynamics introduced in Section III, we consider the scenario illustrated in Figure 5 of Section II, and use our methodology for performing stochastic analysis of the original dtSHS through the Markov Chain abstraction.

We consider a simplified Performance Based Airspace (PBA) as a rectangular airspace defined by $a = 0, b = 10 \text{ km.}, c = 0, d = 10 \text{ km.}$, as illustrated in Figure 8.

The continuous time stochastic dynamics of the aircraft are discretized with sampling time $\Delta = 1 \text{ s}$ using the *Eulero-Maruyama* discretization, and the continuous state space of the aircraft dynamics have been restricted to the simplified PBA and partitioned using a grid of width $\delta/\sqrt{2}$, as illustrated in Figure 8. The parameter δ is the diameter of each partition cell. The number of cells is $n \cdot m$, and depends on a, b, c, d , as follows:

$$m = \frac{(d - c)}{(\delta/\sqrt{2})}, n = \frac{(a - b)}{(\delta/\sqrt{2})}.$$

BADA PERFORMANCE FILE										2000/12/07			
AC/Type: B763_				Last BADA Revision: 3.3		Source OFP File: 3.3		2000/12/06		Source APF File: 3.3		2000/12/06	
Speeds:	CAS (LO/HI)	Mach	Mass Levels [kg]			Temperature: ISA							
climb	- 250/290	0.78	low - 107880										
cruise	- 250/310	0.80	nominal - 150000			Max Alt. [ft]: 43000							
descent	- 250/290	0.78	high - 181400										
FL	TAS [kts]	CRUISE fuel lo [kg/min] nom hi	TAS [kts]	ROCD To [fpm] nom hi	CLIMB fuel kg/min nom	TAS [kts]	ROCD [fpm] nom	DESCENT fuel kg/min nom					
0			164	2230 1970 1650	251.7	152	790	70.0					
5			165	2220 1970 1650	249.4	153	810	69.3					
10			166	2210 1950 1640	247.1	159	850	68.9					
15			172	2320 2030 1710	245.6	171	940	68.7					
20			174	2310 2020 1690	243.3	203	1020	34.7					
30	230	51.1 72.3 92.6	197	2770 2340 1960	241.9	230	1220	34.5					
40	233	51.2 72.5 92.8	231	3360 2750 2310	242.0	233	1250	33.9					
60	272	55.6 72.5 88.7	272	4260 3010 2380	237.6	240	1300	32.6					
80	280	55.8 72.8 89.2	280	4120 2890 2270	228.4	280	1290	16.6					
100	289	56.0 73.2 89.6	289	3970 2770 2160	219.3	289	1700	16.1					
120	297	56.2 73.5 90.1	344	4060 2880 2300	215.8	344	2010	15.5					
140	306	56.4 73.9 90.7	354	3870 2730 2160	206.6	354	2030	15.0					
160	389	72.5 84.7 96.3	365	3670 2570 2010	197.4	365	2050	14.5					
180	401	72.7 85.0 96.8	376	3460 2400 1860	188.2	376	2070	13.9					
200	413	72.8 85.3 97.2	387	3250 2230 1700	179.1	387	2090	13.4					
220	425	72.9 85.6 97.7	399	3030 2050 1540	169.9	399	2100	12.8					
240	438	72.9 85.9 98.2	412	2800 1860 1370	160.7	412	2120	12.3					
260	452	73.0 86.1 98.7	425	2560 1670 1200	151.6	425	2140	11.8					
280	466	73.1 86.4 99.2	438	2310 1470 1020	142.4	438	2150	11.2					
290	473	73.1 86.6 99.5	445	2190 1370 930	137.8	445	2160	10.9					
310	469	69.3 84.0 98.2	458	2710 1630 1040	128.6	458	3040	10.4					
330	465	66.0 82.1 97.5	454	2390 1340 750	118.2	454	2940	9.9					
350	461	63.2 80.8 97.8	450	2050 1040 450	108.0	450	2870	9.3					
370	459	60.9 80.4 96.9	447	1570 660 120	98.0	447	2600	8.8					
390	459	59.4 80.8 87.2	447	1250 360 0	88.2	447	2600	8.2					
410	459	58.3 77.7 77.7	447	920 50 0	78.6	447	2620	7.7					
430	459	57.9 68.2 68.2	447	570 0 0	69.0	447	2670	7.2					

Fig. 7: Data obtained from BADA [11].

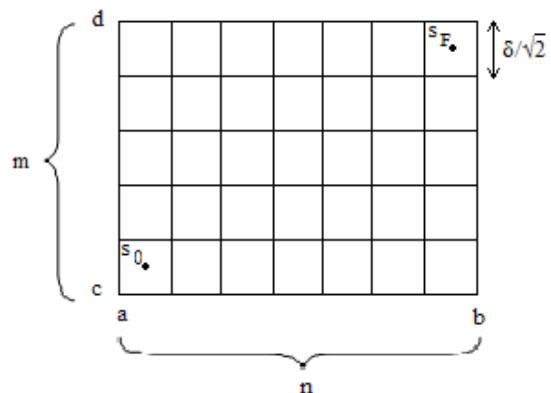


Fig. 8: Simplified and partitioned Performance Based Airspace (PBA).

We have chosen $\delta/\sqrt{2} = 0.2 \text{ km.}$, which generates a 50×50 grid. Using the space and time discretizations defined above, we construct using the abstraction method defined in [9] a Markov Chain abstraction of the original dtSHS. This Markov Chain is defined by a 2501×2501 stochastic matrix Π and an initial probability distribution Π_0 , which in our case is given by the 2501×1 vector $\Pi(0) = [1 \ 0 \ 0 \dots 0]^T$ (i.e. the initial position of the aircraft is the departure TMA with probability 1. The 2501st state of the Markov Chain is an absorbing sink

state, that models the state space region $\mathbb{R}^2 \setminus [a, b] \times [c, d]$. It is reasonable to model this whole region as a single unsafe state, since it models that the aircraft exits the PBA without reaching the arrival TMA. Moreover, the probability of entering this state is usually $\cong 0$.

A. Probability distribution evolution

Given Π, Π_0 , we can compute the stochastic evolution at step $t \in \mathbb{N}$ of the 2501×1 probability vector $\Pi_t = \Pi^t \Pi_0$. As illustrated in Figure 10 we performed computation of Π_t at time steps $t = 1, 2, \dots, T$, using MATLAB for constructing Π, Π_0 and plotting Π_t . The x-y plane represents the PBA, and the z axis is the probability that the aircraft belongs to each cell.

In Figure 9, it is clear that the effect of wind might bring the aircraft in the Weather hazards areas WHA, even if with a small probability. Computing the probability distribution Π_t for $t \in \{1, \dots, T\}$, it is possible to compute the probability of entering the WHA area in the time interval $[0, T\Delta]$. In our case study, we considered $T = 70$.

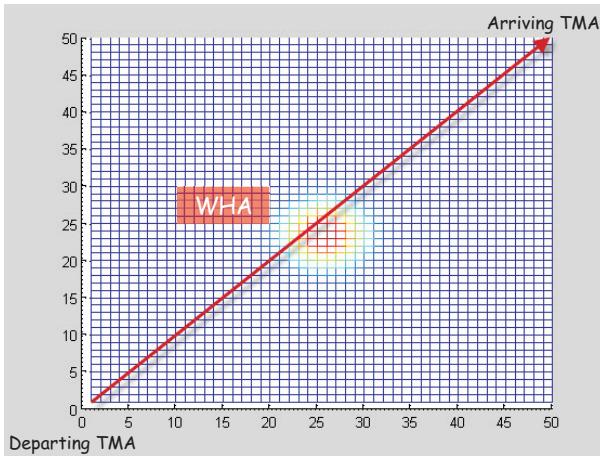


Fig. 9: Trajectory deviation at time 30 s. due to the wind deterministic component.

B. PCTL model checking

Given the Markov Chain abstraction and the WHA area defined by the set $[2, 4]km \times [5, 6]km$, we use the obtained matrix Π as an input to the tool PRISM [5], [6], [7], in order to perform model checking of the following PCTL properties.

1. Does the aircraft eventually reach the arriving TMA point, with probability greater than a value P ? This formula can be expressed in PCTL by the unbounded until formula

$$\text{TRUE} \mathcal{U} \text{TMA}. \quad (4)$$

Using PRISM with our abstraction we verified that, on our model, the property is satisfied with probability $P \geq 0.85$.

2. Does the aircraft eventually reach the arriving TMA point without passing through the WHA area, with probability greater than a value P ? This formula can be expressed in PCTL by the unbounded until formula

$$\text{WHA} \mathcal{U} \text{TMA}. \quad (5)$$

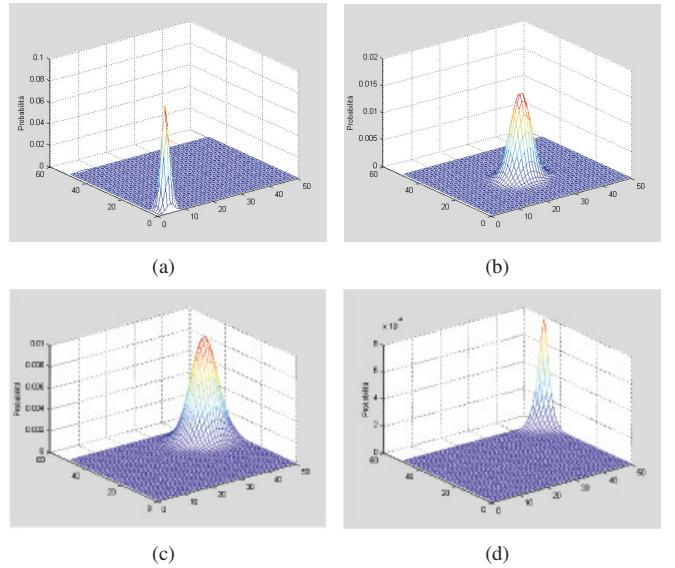


Fig. 10: Probability distribution evolution at times 5 s. (a), 30 s. (b), 50 s. (c), 70 s. (d).

Using PRISM with our abstraction we verified that, on our model, the property is satisfied with probability $P \geq 0.80$.

3. Does the aircraft reach the arriving TMA point within the Controlled Time of Arrival (CTA) and without passing through the WHA area, with probability greater than a value P ? This formula can be expressed in PCTL by the bounded until formula

$$\text{WHA} \mathcal{U}_{\leq \text{CTA}} \text{TMA}. \quad (6)$$

Using PRISM with our abstraction we verified that, on our model, the property is satisfied with probability $P \geq 0.25$ for $\text{CTA} = 50s.$, with probability $P \geq 0.77$ for $\text{CTA} = 60s.$, and with probability $P \geq 0.80$ for $\text{CTA} = 70s..$

The abstraction approximation introduces an error in the probabilistic evolution of the Markov Chain with respect to the original dtSHS, which depends on δ . In this paper, and using the bounds derived in [9], we considered a partition that introduces an error of 0.1 in the steady state probability of the abstraction. This is due to the limited resources of the hardware used for the simulations (a 1.8 GHz CPU takes 1 hour for constructing the Markov Chain abstraction). However, according to the results illustrated in [9], the precision of the abstraction can be arbitrarily chosen by decreasing δ and using faster CPUs. Moreover, executing model checking to our model through the tool PRISM is extremely fast on the abstraction Markov Chain (it takes a few seconds for each PCTL formula) even with a slow CPU. It is fundamental to stress that there exist no tools that perform stochastic model checking over a dtSHS: for this reason, our methodology is a technological enabler for applying automatic stochastic model checking to dtSHSs. Using model checking through our methodology, it is possible to determine the probability that the aircraft reaches the arriving TMA. The crew can use this value to decide whether to continue on the Business Trajectory, or to change the flight plan in order to avoid the WHA. For this reason, our methodology can be a useful tool to validate

and apply new ATM concepts (in our case study, A^3 ConOps).

V. CONCLUSIONS

In this paper, we apply a methodology for formal reasoning based on stochastic hybrid systems theory and abstraction algorithms for dynamical systems, that provides a powerful framework to analyze multi-agents stochastic models of ATM procedures. We use PRISM model checker tools for verifying PCTL probabilistic properties of ATM scenarios. We applied our methodology to a simple single-agent ATM scenario, in the context of the concept A^3 ConOps. Future work aims to apply our methodology in a compositional framework, in order to overtake computational complexity issues in multi-agent systems.

ACKNOWLEDGMENTS

This work was partially supported by European Commission under STREP project n.TREN/07/FP6AE/S07.71574/037180 iFLY.

REFERENCES

- [1] J.-M. Loscos, "Asas:towards new cooperation based on airborne spacing," *Revue Technique de la DTI, ISSN 776-1239*, December 2005.
- [2] "D6.1b qualitative risk assessment for asep-itp, v.1.0," ASSTAR Projects, Tech. Rep., 01 February 2007.
- [3] G. Cuevas, I. Echegoyen, J. Garcia, P. Casek, C. Keinrath, R. Weber, P. Gotthard, F. Bussink, and A. Luuk, "Autonomous Aircraft Advanced A^3 ConOps," iFly, Tech. Rep., January 2009, deliverable 1.3.
- [4] M. Kwiatkowska, G. Norman, and D. Parker, "Stochastic model checking," in *Formal Methods for the Design of Computer, Communication and Software Systems: Performance Evaluation (SFM07)*, ser. Lecture Notes in Computer Science, M. Bernardo and J. Hillston, Eds. Springer, 2007, vol. 4486 (Tutorial Volume), pp. 220–270, to appear.
- [5] A. Hinton, M. Kwiatkowska, G. Norman, and D. Parker, "PRISM: A tool for automatic verification of probabilistic systems," in *Proc. 12th International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS'06)*, ser. LNCS, H. Hermanns and J. Palsberg, Eds., vol. 3920. Springer, 2006, pp. 441–444.
- [6] M. Kwiatkowska, G. Norman, and D. Parker, "Prism: Probabilistic model checking for performance and reliability analysis," *ACM SIGMETRICS Performance Evaluation Review*, 2009.
- [7] "www.prismmodelchecker.org."
- [8] A. Abate, M. Prandini, J. Lygeros, and S. Sastry, "Probabilistic reachability and safety for controlled discrete time stochastic hybrid systems," *Automatica*, 2007, accepted for publication.
- [9] A. D'Innocenzo, A. Abate, M. Di Benedetto, and S. Sastry, "Approximate abstractions of discrete-time controlled stochastic hybrid systems," in *Proceedings of the 47th IEEE Conference of Decision and Control*, Cancun, MX, December 2008, pp. 221–226.
- [10] W. Glover and J. Lygeros, "Deliverable number d1.3: A multi-aircraft model for conflict detection and resolution algorithm evaluation," HYBRIDGE, Tech. Rep., February 18 2004, project: Distributed Control and Stochastic Analysis of Hybrid Systems Supporting Safety Critical Real-Time Systems Design.
- [11] "www.eurocontrol.int/eec/public/standard_page/proj_bada.html."

Doctoral Symposium

Establishing an Upper-Bound for the Benefits of NextGen Trajectory-Based Operations

Guillermo Calderón-Meza (Ph.D. Candidate)
 Research Assistant
 Center for Air Transportation Systems Research
 George Mason University
 Fairfax, Virginia, 22030, USA
 Email: gcaldero@gmu.edu

Lance Sherry (Ph.D.)
 Director
 Center for Air Transportation Systems Research
 George Mason University
 Fairfax, Virginia, 22030, USA
 Email: lsherry@gmu.edu

Abstract—NextGen enabling technologies and operational initiatives seek to increase the effective-capacity of the National Airspace System. Concepts-of-operations, such as Trajectory-Based Operations, will allow flights increased flexibility in their 4-D trajectories as they traverse Center airspace. Shifting trajectories in this way can minimize the airlines operating costs (i.e., distance flown), shift the geography of Air Traffic Control (ATC) workload (i.e., sectors used), shift the time-intensity of ATC workload (i.e., flights counts per sector).

This paper describes the results of an analysis of one day of operations in the NAS using traditional navigation aid-based airway routes compared to direct, i.e., Great Circle Distance, routes. The results yield: (i) a total of 599 thousand nm (average 30 nm per flight) savings generated by flying direct routes, (ii) a redistribution of flights across sectors resulting in a reduction of 3% in the total time the flights in a sector are in excess of the Monitor Alert Parameters for that sector, (iii) a reduction in ATC workload reflected by a 47% drop in the number of flights requiring conflict resolution. These results indicate upper bound of benefit opportunities for both ATC and the airlines based on the introduction of flexible routing structures in NextGen.

Index Terms—NextGen, evaluation, conflicts, FACET, distance flown, delays.

I. INTRODUCTION

NextGen [1] enabling technologies and operational initiatives seek to increase the effective-capacity of the *National Airspace System* (NAS) by opening up unused airspace, increasing the availability of airspace in all weather conditions, and increasing the utilization of existing airspace by reducing spacing between flights on the same routes.

Concepts of operations, such as *Trajectory-based Operations* (TBO), will allow flights increased flexibility in their 4-D trajectories as they traverse center airspace. Whereas the airlines may benefit from reduced distance flown, the adjustment of the 4-D trajectories will shift the geographic distribution of flights across *Air Traffic Control* (ATC) sectors, as well as the distribution of instantaneous flight counts in individual sectors. Several related concepts are identified as Trajectory-Based Operations [1], [2]: 1) *Continuous Descent Arrivals* (CDA) that smooth the transition from top-of-decent to near idle speed. These include *Tailored Arrivals* that use technology (automation tools and data communications) to

provide a preferred trajectory path and transfer it to the flight management system on the aircraft. 2) *3D Path Arrival Management* that designs fuel-efficient routes to decrease controller and pilot workloads. 3) *4D Trajectory-Based Management* that defines 3-dimensional flight paths based on points in time (the 4-D) from gate-to-gate. 4) *Required Navigation Performance* in which navigation performance requirements for operation within an airspace define the trajectories. In this paper, only the third definition is considered.

This paper analyzes the potential upper bound of impact of the shifting trajectories to minimize the airlines operating costs (i.e., distance flown), the geographic workload (i.e., sectors used), and the time workload (i.e., flights counts per sector) for Air Traffic Control (ATC). Similar studies have been carried out to evaluate the impact of this change and other changes proposed by NextGen. Barnett [3] evaluates the impact in safety caused by using direct routes instead of airways. The study concludes that using direct routes diminishes the risk of en-route collision. These results are valid only if certain rules for TFM remain in effect after the change. A caveat of the study is that the results will depend on the capacity of the technology and humans to match the current performance of the ATC. Agogino and Turner [4] evaluate policies intended to optimize performance of the TFM. The metrics used are congestion and delays. The study evaluates several ATC algorithms as well as the use of multi-agent technology. The algorithms achieve significant improvements in performance compared to previous algorithms and the current practices. Magill [5] also analyzes the change from airway routes to direct routes. The study uses the number of conflicts (called interactions in this case) as an approximate metric of the ATC workload. The study modifies the separation rules as well as the type of routes. The paper concludes that the reduction of traffic density due to the use of direct routes is the most significant factor in the reduction of workload for the ATC.

The Future ATM Concept Evaluation Tool (FACET¹) [6] was used for this experiment that included 19,900 domestic flights between 287 airports (4,170 O/D pairs). The experiment

¹See www.aviationsystemsdivision.arc.nasa.gov/research/modeling/facet.shtml

consisted of two scenarios: (i) flights followed *Great Circle Distance* (GCD) routes from TRACON to TRACON, and (ii) flights followed traditional navigation aid-based airway routes. The flights in each scenario used the same cruise flight levels and cruise speeds. The results are summarized below:

- (i) Great Circle Distance routes generated a total of 598,724.8 nm (average 30.1 nm per flight) savings in reduced distance flown.
- (ii) Great Circle Distance routes resulted in a redistribution of ATC workload reducing the time sectors were above their *Monitor Alert Threshold* (MAP) from 32% to 21%.
- (iii) Great Circle Distance routes resulted in reduced ATC workload reducing the number of flights with conflicting trajectories by 47%.

These results establish an upper bound on the benefits to be derived by Trajectory-based Operations. The result is a win-win scenario for both the airlines and air traffic control. The use of Great Circle Distance routes geographically redistributed the flights reducing workload in the most congested sectors and well as significantly reducing conflicts in flight trajectories. It should also be noted that the use of Great Circle Distance routes did not alleviate the flight delays resulting from over-scheduled departure and arrivals.

This paper is organized as follows: Section 2 describes the design of the experiment, the simulation used for the experiment, and the configuration and parameters used in the experiment, Section 3 describes the results of the experiment, and Section 4 provides conclusions, implications of these results, and future work.

II. METHOD AND DESIGN OF THE EXPERIMENT

The experiment was conducted using the Future ATM Concept Evaluation Tool (FACET) [6]. The tool has been used in previous studies [4], [7], [8] to evaluate new *Traffic Flow Management* (TFM) concepts in the NAS. FACET offers many options like the possibility connecting to real-time data sources for weather and traffic, real-time conflict detection and resolution, batch processing of input data (as an option to real-time streams), and a Java API². In the absences of random inputs (like weather phenomena) the simulation is deterministic. The results will be the same regardless of the number of executions.

Other metrics of the system, like number of sectors or centers flown, distance flown, and number of conflicts, can be obtained from the API or from the GUI³.

A. The input files for FACET

The main input to FACET is the flight schedule, flight tracks and cruise flight-levels. FACET accepts several formats for these input files known as ASDI, TRX. To achieve the goals of this experiment, a TRX input file was generated based on actual historical data from the *Airline On Time Performance Data* data provided by *Bureau of Transportation Statistics*

²API: Application Program Interface.

³GUI: Graphical User Interface.

(BTS). The procedure for generating the TRX file is described bellow.

First, the sample TRX files that come with FACET were parsed and the O/D pairs and corresponding flight plans were extracted and exported to a database.

Second, the BTS Airline On-Time Performance (AOTP) data was queried to obtain a single day of domestic operations. The query extracted the O/D pair, the coordinates for the airports (taken from a proprietary table), the scheduled departure and arrival times, the flight and tail numbers, and the aircraft type (taken from a proprietary table related to On Time by tail number). The results of this query are sorted, ascending, by scheduled departure time.

For each record returned by the query the great circle distance of the O/D pair, the expected flight time (that is the difference of the scheduled departure and arrival times both converted to GMT), the required ground speed (and integer number of knots), the heading (an integer number computed from the coordinates of the airports assuming 0 degrees for North heading, and 90 degrees for West heading), and the flight level (a uniformly distributed random integer number from 200 to 450), and the flight plan (taken randomly from available plans for the O/D pair). The coordinates of the airports are converted into integer numbers with the format [+|-]DMS where D stands for degrees (two or three digits), M stands for minutes (two digits), and S stands for seconds (two digits). FACET requires western longitudes to be negative.

Third, for each group of records with the same GMT scheduled departure time one "TRACKTIME" record is written to a text file. The value of the TRACKTIME record is the GMT scheduled departure date/time converted into the number of seconds from January 1, 1970 GMT. After this TRACKTIME record, the individual "TRACK" records for the flights are written using the data computed in the second step. The process repeats until there are no more records from the query. An input file generated this way does not track the flights through the National Airspace System. It only describes every flight with a single record. So this file can be used for simulation purposes only, not for playback in FACET.

The file used in this experiment contains 19,900 domestic (USA) flights scheduled to departure from Friday July 27 2007 at 05:30:00 GMT to Saturday July 28 2007 at 09:20:00 GMT. The actual landing date/time of the last flight differs between scenarios because flights could be delayed or they could fly different distances.

B. Design of Experiment

The goal of this experiment is to evaluate the effect of changing from the current airway routes, i.e., flight plans, to Great Circle Distance routes, i.e., direct routes, as it is proposed by NextGen.

This paper presents and compares the results of one experiment divided into two scenarios (see Table IV). The first scenario simulates one day of NAS operations in which all the flights use airway routes, i.e., flight plans, as it is done today in the NAS. The second scenario simulates the same day of

operations, but flights follow Great Circle Distances routes, i.e., direct routes, between the origin and the destination.

The outcomes of interest for each scenario are the total number of *centers* and *sectors*, and the *distance* flown by the flights, the total number of *conflicts* detected, and the *flight delays* generated in the OEP-35⁴ airports. The benefits and costs for the airlines, controllers, and the environment can be computed using these outcomes.

The distances flown are compared using a paired two-tail t-test. The paired t-test applies since the simulator (FACET) uses the same input file in both scenarios, so each flight in one scenario can be compared to its similar in the other scenario. However, it was observed that some flights do not appear in both scenarios, even if the input file is the same. The reasons for this fact are still not completely understood. But, only flights that appear in both scenarios are used in the t-test.

The comparison of the total number of conflicts is only done for a single pair of numbers, so no statistical test is applied in this case.

The distribution of the sectors load is multi-dimensional. There is spatial distribution and temporal distribution. In this paper mainly the temporal distribution will be analyzed, leaving the spatial distribution for future work. The two scenarios are compared using the percentage of time in which at least one sector contains a number of flights that is on or over the sectors Monitor Alert Parameter (MAP) value, i.e., it is overloaded. To get an idea of the distribution, also the percentage of time in which at least one sector is at or over 80% of its MAP is compared between scenarios.

No external disturbances are included during the simulations, i.e., there are no restrictions due to weather, congestion, push-back delays, or other stochastic events. So, the simulations are deterministic. The only limitation that is imposed in the arrival capacity of the EOP-35 airports, which is set to the VFR departure and arrival rates for the whole day (see Table I). Even with VFR rates, this limitation generates ground delays via *Ground Delay Programs* (GDP), but their effect is not strong because the airports are not significantly over-scheduled in the scenario input file used for the experiment. However, the ground delays are compared using the total minutes of delays and the average minutes in the OEP-35 airports which are the only ones restricted using GDPs.

C. FACET settings used in experiment

In this experiment, FACET takes its input from batch files, and the outputs are taken from the simulation via the API. The input file was loaded using the *loadDirectRouteSimAsynch* and the *loadFlightPlanSimAsynch* functions of the API. With the first function, FACET sets itself to use Great Circle Distance routes, i.e., direct routes. With the second function, FACET uses the airways routes, i.e., flight plans, provided in the input file. Both functions accept the same number and types of arguments. The *trajectory update interval* is set to 60. The *integration time step* is set to 60.0. And the *additional update delay* is set to 0.2.

⁴OEP: Operational Evolution Partnership Plan.

TABLE I
DEFAULT VFR AIRPORT ARRIVAL RATES (AAR) FOR THE OEP-35 AIRPORTS USED IN THE SIMULATION

Airport name (ICAO)	Airport Arrival Rate (Moves per hour)	Airport name (ICAO)	Airport Arrival Rate (Moves per hour)
KATL	80	KLGA	40
KBOS	60	KMCO	52
KBWI	40	KMDW	32
KCLE	40	KMEM	80
KCLT	60	KMIA	68
KCVG	72	KMSP	52
KDCA	44	KORD	80
KDEN	120	KPDX	36
KDFW	120	KPHL	52
KDTW	60	KPHX	72
KEWR	40	KPIT	80
KFLL	44	KSAN	28
KHNL	40	KSEA	36
KAID	64	KIAH	72
KSFO	60	KJFK	44
KSLC	44	KLAS	52
KSTL	52	KLAX	84
KTPA	28		

The API provides an interface (*ConflictInterface*) with functions to enable (*setEnabled*) and configure (*setConflictDetectionParameters*) the conflict detection functionality. The parameters are as follows. The center index is set to -1, i.e., all the centers. The *surveillance zone* is 120 nm. The lookahead time is 0. The *horizontal separation* is 6 nm. The *vertical separation* below f1290 is 1000 ft. The *vertical separation* above f1290 is 1000 ft. Also, the detected conflicts are displayed during the simulation.

The arrival rates (AAR) of the airports are infinite by default in FACET. For this experiment, the capacities are limited using FACET's GDP functionality. The OEP-35 airports are assigned a maximum capacity in the form of an *arrival delay GDP* from the 0:00 to 24:00 (see Table I). With these limits in the capacity of the airports, FACET starts recording delay statistics during the simulations.

A total of 35 hours and 55 minutes (2,155 minutes) of operations are simulated. All other parameters of FACET are left to their default values.

An external Java program, using the FACET API, measures the distance traveled as follows. At each simulation time step (one minute) the distance flown by each flight is updated based on the previous and current coordinates. The computation of distance is done with the *utils.getGCDistanceNM* function of FACET's API. The external program also records the total number of flights in each sector, including all the sector levels, i.e., low, high, and super. Distances and sector loads are written into text files for further analyses.

III. RESULTS

A. Distances flown

Figure 1 compares the histogram of the distance flown when using airways to the histogram of the distance flown when using direct routes. The figure also includes the descriptive statistics for the scenarios. When using airways, most of

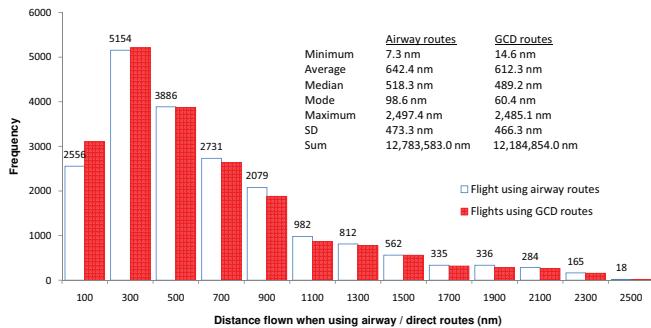


Fig. 1. Comparison of histograms of the distances flown using airways and direct routes

the flights travel less than 1,000 nm, with a peak of flights between 200 and 400 nm. Short flights, i.e., less than 200 nm are frequent, but not a majority in this input file. There is a long tail in the distribution, but the actual number of flights is low compared to the other distances. The flights with longer distances correspond to flights from Alaska or other US territories not directly in the continent.

When using direct routes, most of the flights travel less than 1,000 nm, with a peak of flights between 200 and 400 nm. Short flights, i.e., less than 200 nm are more frequent than when using airways. This is an immediate benefit of using direct routes: shorter flown distances. The comparison of the tails shows that their frequencies are similar.

The distribution for the scenario of the direct routes is shifted toward the shorter distances. This is evident in that the average, median, and mode are smaller in this scenario than they are in the airways scenario. The standard deviation is also smaller indicating that the distribution is less disperse in this scenario.

The figure shows that the input file used in this experiments is dominated by short to mid distance distance flights. This reflects that the input file comes from a database that contains only data for actual domestic flights in the US. The greater changes in the frequencies are observed in the flights from 0 to 200 nm, and from 800 to 1,000 nm. This suggests that the benefits of using direct routes are clearer in short flights or in trans-continental flights.

Figure 2 shows the distribution of the differences of distance flown by corresponding flights in both scenarios, i.e., it is a paired comparison of distances. The figure also includes the descriptive statistics for the distribution. The 1,093 (5.5%) of the differences in the distance flown are negative indicating that the direct routes are longer than the airway routes. This is mathematically incorrect. This is due to errors in the measurement of the distance during the simulation. Notice that the minimum difference is -7.9 nm, and the bin of the histogram goes from -100 to 0 nm, so the negative differences are in this 7.9 nm range. The peak of the histogram occurs when the difference is between 0 and 100 nm, 90% of the differences are in this range.

A paired two-tail t-test shows that the mean of the difference

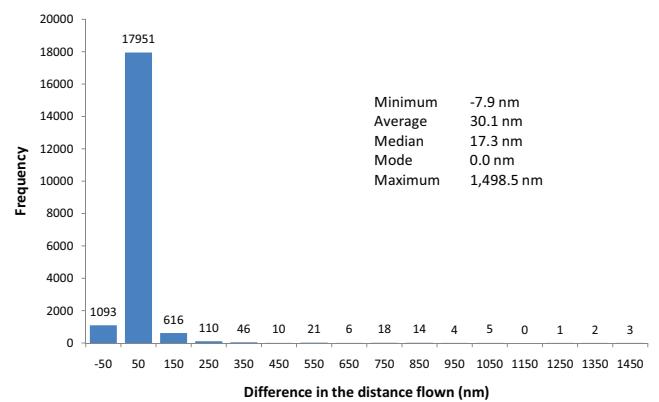


Fig. 2. Histogram of the flight-by-flight difference in the distance flown

TABLE II
NUMBER OF MINUTES WITH AT LEAST ONE SECTOR SATURATED OR ON THE VERGE OF SATURATION

Scenario	Number of minutes with at least one sector on or above (% of the total 2,114 minutes)	
	MAP	80% of MAP
Flights using airway routes	689 (32%)	944 (44%)
Flights using direct routes	456 (21%)	917 (43%)

between the distances flown by corresponding flights in the two scenarios is significantly different than zero ($M=30.1$, $SD = 60.7$, $N = 19,900$), $t = 69.9647$ and the two-tail $p = 0.000$. A 95% confidence interval about the mean is (29.2, 30.9). This average reduction in the distance flown adds to 598,724.8 nm saved when using direct routes instead of airways. The reduction in distance flown benefits the airlines and the environment, through a reduction in fuel burned, i.e., less pollution and lower costs.

B. Sectors over MAP

A metric for the load of sectors is a function of time, space, the number of flights, and routes of the flights. The number of flights did not change between scenarios in this experiment. The routes are expected to change significantly when going from flight plans to direct routes. With this change in the type of route the distribution of sector load through time and space is also expected to change.

The time distribution of the sector load is analyzed in this experiment. TABLE II shows that controllers spend 32% of their time managing congested sectors, i.e., at or above the sector's MAP, when the flights use airway routes. But controllers spend 21% of their time managing congested sectors when the flights use direct routes. The values for 80% of MAP give an idea of the distribution of sector load in the two scenarios. The percentage of time controllers spend managing sectors with 80% or more of their MAPs is similar in both scenarios with a small reduction for the case of direct routes. This similarity indicates that using direct routes mostly reduces the frequency of overloaded sectors, but does not change the total time controllers spend managing "almost saturated" sectors.

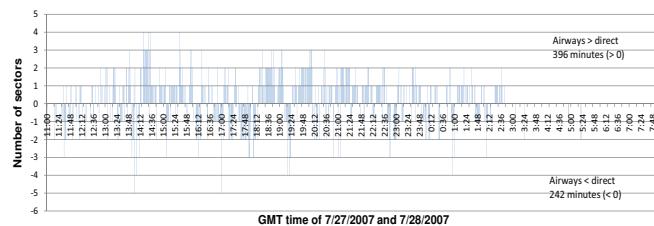


Fig. 3. Minute by minute difference in the number of sectors on or above MAP

Comparing the sector loads minute by minute provides more insight of effect of using direct routes in the NAS. Figure 3 shows that using airways produces load peaks (396 minutes, positive side of the vertical axis) that are often higher in value and closer in time than when using direct routes. Using direct routes produce few intense peaks (value -4 and -5), but the peaks (242 minutes) are more scattered in time. So controllers will have more time to “rest” between peaks of saturation when flights use direct routes and the saturation will be, in average smaller than when using airways.

C. Conflicts

The total number of conflicts detected reduced from 23,071 when using flight plans to 12,308 when using direct routes. This is an improvement in safety, i.e., lower probability of accident, and a further reduction in the workload of the controllers, i.e., they have to resolve 46.7% less conflicts. Magill [5] found that, for similar separation rules, the reduction was about 35%.

D. Delays

The flight ground delays generated by the GDPs defined for the OEP-35 airports are summarized in Table III. The arrival capacities of the OEP-35 airports were set to VFR rates for the whole day. Ground Delay Programs were activated at all the OEP-35 airports.

The total ground delay generated for the OEP-35 airports reduces from 14,076.4 minutes when using airway routes to 13,444.0 minutes when using direct routes. The average delay for all the OEP-35 airports remains similar between scenarios: the reduction is in the order of few seconds.

The mean flight delay differs from airport to airport ranging from 7.5 min to 0.3 min in the case of the airway routes, but from 6.8 min to 0.3 min in the case of direct routes. These numbers are low with respect to the observations of the actual airports due to (i) absence of international flights, (ii) the scenarios resulted in the same degree of over-scheduling of departures and arrivals. The effect of the direct routing would be equally likely to over-schedule arrivals as it would be to reduce simultaneous arrivals.

TABLE IV summarizes the results of the two scenarios and the previous tables and charts.

IV. CONCLUSIONS

This experiment consisted of two scenarios with the same set of 19,900 domestic flights in the NAS. The scenarios were

TABLE III
FLIGHT DELAYS ON THE OEP-35 AIRPORTS OBTAINED BY LIMITING ARRIVAL CAPACITY

Airport code (ICAO)	Flight plan			Direct route		
	Number of flights	Total delay (min)	Avg delay (min)	Number of flights	Total delay (min)	Avg delay (min)
KATL	941	7,066.9	7.5	960	6,550.3	6.8
KBOS	167	90.0	0.5	169	85.0	0.5
KBWI	147	120.8	0.8	148	143.5	1.0
KCLE	130	188.8	1.4	132	206.2	1.6
KCLT	262	275.7	1.0	265	265.7	1.0
KCVG	244	246.5	1.0	247	234.8	1.0
KDCA	161	145.0	0.9	163	156.1	1.0
KDEN	413	128.2	0.3	416	131.2	0.3
KDFW	562	178.3	0.3	569	188.7	0.3
KDTW	203	163.2	0.8	206	167.0	0.8
KEWR	212	278.7	1.3	213	246.5	1.2
KFLL	103	84.4	0.8	105	90.0	0.8
KHNL	81	85.0	1.0	88	54.8	0.6
KIAD	117	74.9	0.6	117	114.6	1.0
KIAH	363	278.8	0.8	364	252.5	0.7
KJFK	156	168.6	1.1	157	178.4	1.1
KLAS	216	207.0	1.0	222	249.3	1.1
KLAX	328	173.4	0.5	333	167.6	0.5
KLGA	187	207.0	1.1	191	221.8	1.2
KMCO	219	191.6	0.9	222	179.2	0.8
KMDW	163	315.5	1.9	164	241.6	1.5
KMEM	93	35.7	0.4	95	43.7	0.4
KMIA	111	55.4	0.5	113	53.5	0.5
KMSP	247	330.3	1.3	249	277.4	1.1
KORD	695	885.2	1.3	706	921.8	1.3
KPDX	83	84.3	1.0	84	81.3	1.0
KPHL	183	161.2	0.9	186	166.0	0.9
KPHX	76	30.3	0.4	76	28.2	0.4
KPIT	66	23.2	0.4	66	20.2	0.3
KSAN	128	244.0	1.9	130	248.2	1.9
KSEA	139	206.3	1.5	142	170.2	1.2
KSFO	179	98.0	0.5	184	91.8	0.5
KSLC	267	903.7	3.4	268	894.8	3.3
KSTL	119	93.1	0.8	121	89.0	0.7
KTPA	136	257.5	1.9	137	233.0	1.7
Totals	7,897	14,076.4	1.8	8,008	13,444.0	1.7

TABLE IV
EXPERIMENTAL DESIGN AND EXPERIMENTAL RESULTS

	Scenario	
	Great Circle Distance routes	Airway routes
Total distance flown (Average distance per flight)	12,184,854.0nm (612.3nm)	12,783,583.0nm (642.4nm)
Percentage of time with sectors above MAP threshold (% of time with sectors 80% or more of MAP)	21% (43%)	32% (44%)
Number of airborne conflicts detected by ATC	12,308	23,071
Total flight delays (Average delays)	13,444.0min (1.7min)	14,076.4min (1.8min)

executed using FACET. In one scenario flights used airways the same way they currently do in the NAS. In the second scenario flights used direct routes. The arrival rate of the OEP-35 airports was set to the VFR rates using the GDP functionality provided by FACET.

The goal of the experiment was to evaluate the effect of introducing direct routes for domestic flights.

The distance flown is smaller, in average 30.1 nm, when flights use direct routes. And the difference is statistically significant. There are more flights with routes of less than 200 nm when flight use direct routes than when they use airway routes. But all the other route distances are less frequent in the case of direct routes than in the case of airways. This reduction in the distance flown results in savings of fuel and time. Airlines and the environment benefit from such a reduction.

Sector congestion is also reduced by using direct routes instead of airway routes. Controllers spend 21% of their time managing overloaded sectors when the flights use direct routes, but they spend 32% when flights use airways.

Peaks of sector congestion are also more separated in time. This reduction might result in safety benefits.

The total number of conflicts detected is reduced about 46.7% (from 23,071 to 12,308) when using direct routes. This results in safety benefits by a reduction of the workload of the controllers.

Ground delays (at the origin airports) reduced when using direct routes, but the reduction is not significant. There was a limitation in the way FACET uses to assign delays that did not allow, in this experiment, to measure the airborne or arrival delays. The delays recorded are only due to the GDPs. And the GDPs are using maximum arrival rates for the OEP airports. This does not impose enough restrictions and generates small delays.

Implications of results

These results establish an upper bound on the benefits to be derived by Trajectory-based Operations. The result is a win-win scenario for both the airlines and air traffic control. The use of Great Circle Distance routes geographically redistributed the flights reducing workload in the most congested sectors and well as significantly reducing conflicts in flight trajectories. It should also be noted that the use of Great Circle Distance routes did not alleviate the flight delays resulting from over-scheduled departure and arrivals.

Future work

Further work is required to monetize the benefits. For example, how does the reduction in conflicts compares to the reduction in distance in terms of costs? What will be the effect of the distance reduction at the destination airports, e.g. will it produce more congestion?. Also studies with more realistic input files, i.e., including all domestic and international flights, are required to observe congestion and reflect the actual effect of the change. Future work also includes resolution of several anomalies in the results including: (i) great circle distance routes in excess of the associated flight plan routes, (ii) excessive route distance, (iii) missing flights. Detailed statistical data are needed for the delays, e.g., standard deviations, modes, medians, and ranges. More studies in which the conditions of the airports are set to IFR instead of VFR will bring more insight of the problem. The environmental effects of

the reduction in distance must also be studied by using more specific tools.

ACKNOWLEDGMENT

The authors would like to acknowledge the contributions and help from the following persons and institutions.

The research this study is part of is funded by NASA (NRA NNA07CN32A). Furthermore, Natalia Alexandrov, Kapil Sheth, María Consiglio, Brian Baxley, Kurt Neitzke, and Shon Grabbe, all NASA employees, have provided suggestions and comments throughout the whole research process.

From Sensis Corporation, George Hunter and Huina Gao. From George Mason University, Dr. Thomas Speller Jr., Dr. Kenneth De Jong, Dr. Robert Axtell, Dr. George Donohue, Dr. John Shortle, Dr. Rajesh Ganesan, John Ferguson, and Keith Sullivan. From Metron Aviation, Jason Burke, Dr. Terry Thompson, and Norm Fujisaki. From FAA, Joe Post and Tony Diana. They have contributed to the improvement of the research.

Finally, thanks to the Ministerio de Ciencia y Tecnología (Minister of Science and Technology) of Costa Rica.

REFERENCES

- [1] JPDO, *Concept of Operations for the Next Generation Air Transportation System, Version 2.0.* Washington DC, USA: Joint Planning and Development Office, June 2007.
- [2] G. Hayman. (2009, June) Trajectory based operations. [Online]. Available: <http://www.afceaboston.com/documents/events/cnsatm2009/Briefings/Tuesday%20Briefs/Tuesday%20Afternoon/Track%202/20TB0%20PresentationGeneHayman.pdf>
- [3] A. Barnett, "Free-flight and en route air safety: A first-order analysis," in *Operations Research*. INFORMS, Nov-Dec 2000, vol. 48, no. 6, pp. 833–845. [Online]. Available: <http://www.jstor.org/stable/222992>
- [4] A. Agogino and K. Turner, "Regulating air traffic flow with coupled agents," in *Proceedings of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, M. Padgham, Parkes and Parsons, Eds., Estoril, Portugal, May 2008, pp. 535–542.
- [5] S. Magill, "Effect of direct routing on air traffic control capacity," in *Air Transportation Systems Engineering*, G. L. Donohue and A. G. Zellweger, Eds., vol. 193. USA: American Institute of Aeronautics and Astronautics, 2001, pp. 385–396, ISBN 1-56347-474-3.
- [6] K. Bilmoria, B. Sridhar, G. B. Chatterji, K. Sheth, and S. Grabbe, "Facet: Future atm concepts evaluation tool," *Air Traffic Control Quarterly*, vol. 9, no. 1, pp. 1–20, 2001.
- [7] B. Sridhar, G. B. Chatterji, S. Grabbe, and K. Sheth, "Integration of traffic flow management decisions," *American Institute of Aeronautics and Astronautics*, 2002.
- [8] R. Jakobovits, P. Kopardekar, J. Burke, and R. Hoffman, "Algorithms for managing sector congestion using the airspace restriction planner," *ATM*. ATM, 2007.

The Research of Multi-Airport Ground Holding Problem Based on the Schedule Optimization

Ye Bojia

College of Civil Aviation

Nanjing University of Aeronautics and Astronautics,
Nanjing, China
yebojia2005@hotmail.com

Abstract—In this article, a Multi-objective and Multi-airport Optimizing Model (MMOM) is firstly built based on flights schedule, integrating airspace capacities and aircrafts turnover. Then the non-dominated sorting genetic algorithm II (NSGA-II) Algorithm is introduced to propose near-optimal solutions which achieve different goals with different estimated departure times. Finally, an instance concerned the recently schedule of three main airports in China is analyzed in detail, and the simulation results show that the model's correct and efficient.

Keywords-Air traffic flow management; multi-objective; multi-airports; NSGA-II Algorithm

I. INTRODUCTION

In recent years, the multi-airport ground holding problem (MAGHP) has becoming an important research area in air traffic flow management (ATFM). The main strategy of MAGHP is to address the problem of reducing congestion at major airports by means of ground holding, by imposing a ground delay to selected aircraft in order to reduce and possibly to avoid airborne delays and congestion.

The GDP was first systematically described by Odoni (1987). Following this, several deterministic models were developed, in which time varying airport capacities were assumed to be fully known beforehand. Deterministic optimization, formulated as integer program (IP), for MAGHP was first proposed by Vrans [1] et al (1994). Bertsimas and Stock [2] (1998) provided a stronger formulation to the deterministic MAGHP. Navazio and Romanin [4] proposed a deterministic optimization model for MAGHP with banking constraints.

There is a substantial literature [3] [5] [6] on optimization models to support MAGHP. However, in almost all such models, the only objective is to minimize system-wide delays cost, which has two components-ground and airborne delays [7][10]. Also most of the optimization models in the literature are formulated as linear and/or integer programming models, so exorbitant computing times is a common shortcoming under realistic case.

In this paper, in light of Chinese air traffic control situation, the ATFM problem is not only investigated from the point of view of the airlines, but also from different air traffic control departments. So the minimization system-wide delays cost is not enough to give sufficient support in China. Thus, a Multi-

Hu Minghua

College of Civil Aviation

Nanjing University of Aeronautics and Astronautics,
Nanjing, China

objective and Multi-airport Optimizing Model (MMOM) is built based on flights schedule, integrating airspace capacities and aircrafts turnover. Then the NSGA-II Algorithm is introduced to propose near-optimal solutions which achieve different goals with different estimated departure times. Finally, an instance concerned the recently schedule of three main airports in China is analyzed in detail, and the simulation results show that the model's correct and efficient.

II. THE MULTI-OBJECTIVE AND MULTI-AIRPORT OPTIMIZING MODEL FORMULATION

A. Notation

Consider a set of airports $K = \{1, \dots, K\}$ and an ordered set of time periods $\Gamma = \{1, \dots, \Gamma + 1\}$. For instance, K might be the set of 3 busiest Chinese airports, and Γ might be a set of 144 time periods of 5 minutes each, amounting to a time horizon of 12 hours. Φ is the set of all flights of interest, e.g., all flights departing from an airport in K or arriving to an airport in K . In light of Chinese situation, the flight set Φ corresponds to an open network of airports, the flight departures from A will be assigned ground holding delay to adjust its schedule while arrivals in A from the external world is assigned airborne delay.

All notations for MMOM:

A the set of airports

Γ the set of time periods

Φ the set of flights arrive in or departure from A

K the set of continued flights where $K \subset \Phi$

M_t^q the capacity of airport $q \in A$ in time period t

Arr_f the scheduled arrival time of flight f , which is also the earliest time flight f can arrive at its destination airport

Dep_f the scheduled departure time of flight f , which is also the earliest time flight f can depart its origin airport

S_f turnaround time required to refuel, reload, and clean the flight $f \in K$

C_g cost of holding a flight on the ground for one unit of time

C_a average cost of flying an aircraft for one unit of time

The decision variables are:

$$X_{f,t}^q = \begin{cases} 1 & \text{if flight } f \text{ takes off from airport } q \text{ at time } t \\ 0 & \text{otherwise} \end{cases};$$

$$Y_{f,t}^q = \begin{cases} 1 & \text{if flight } f \text{ arrives in airport } q \text{ at time } t \\ 0 & \text{otherwise} \end{cases};$$

$$t \in \Gamma, q \in A, f \in \Phi;$$

The auxiliary variables correspond to multi-objectives are:

$$GA_f = \begin{cases} 1 & \text{if flight } f \text{ assigned ground delay} \\ 0 & \text{otherwise} \end{cases}, f \in \Phi;$$

$$AA_f = \begin{cases} 1 & \text{if flight } f \text{ assigned airborne holding} \\ 0 & \text{otherwise} \end{cases}, f \in \Phi;$$

$$GD_f = \begin{cases} 1 & \text{if ground delay time of flight } f > 15 \text{ mins} \\ 0 & \text{otherwise} \end{cases}, f \in \Phi;$$

$$AD_f = \begin{cases} 1 & \text{if airborne delay time of flight } f > 15 \text{ mins} \\ 0 & \text{otherwise} \end{cases}, f \in \Phi;$$

B. Objective Functions

The multi-objective functions are as follows:

1) Minimize the expected sum of ground and airborne delay time for all flights:

$$\min \sum_{q \in A} \sum_{f \in \Phi} \sum_{t \in \Gamma} [(t - Dep_f) \times (X_{f,t}^q - X_{f,t-1}^q) + (t - Arr_f) \times (Y_{f,t}^q - Y_{f,t-1}^q)]$$

2) Minimize the expected sum of ground and airborne delay costs for all flights:

$$\min \sum_{q \in A} \sum_{f \in \Phi} \sum_{t \in \Gamma} [(t - Dep_f) \times (X_{f,t}^q - X_{f,t-1}^q) \times C_g + (t - Arr_f) \times (Y_{f,t}^q - Y_{f,t-1}^q) \times C_a]$$

3) Minimize the expected sum of adjusted flights:

$$\min \sum_{q \in A} \sum_{f \in \Phi} \sum_{t \in \Gamma} [GA_f \times (X_{f,t}^q - X_{f,t-1}^q) + AA_f \times (Y_{f,t}^q - Y_{f,t-1}^q)]$$

4) Minimize the expected sum of delay flights (delay time > 15mins):

$$\min \sum_{q \in A} \sum_{f \in \Phi} \sum_{t \in \Gamma} [GD_f \times (X_{f,t}^q - X_{f,t-1}^q) + AD_f \times (Y_{f,t}^q - Y_{f,t-1}^q)]$$

C. Constraints

The set of constraints are given by the expressions (1)-(6) below.

$$X_{f,t}^q - X_{f,t-1}^q \geq 0 \quad \forall f \in \Phi, q \in A, t \in \{Dep_f \dots T+1\}; \quad (1)$$

$$Y_{f,t}^q - Y_{f,t-1}^q \geq 0 \quad \forall f \in \Phi, q \in A, t \in \{Arr_f \dots T+1\}; \quad (2)$$

$$\sum_{f \in K} t(X_{f,t}^q - X_{f,t-1}^q) - Arr_f - S_f \leq 0 \quad \forall f \in K, q \in A, t \in \{Arr_f \dots T+1\}; \quad (3)$$

$$\sum_{f \in K} t(Y_{f,t}^q - Y_{f,t-1}^q) - Dep_f - S_f \leq 0 \quad \forall f \in K, q \in A, t \in \{Dep_f \dots T+1\}; \quad (4)$$

$$\sum_{f \in \Phi} X_{f,t}^q + \sum_{f \in \Phi} Y_{f,t}^q \leq M_t^q \quad \forall f \in \Phi, q \in A, t \in \{1 \dots T+1\}; \quad (5)$$

$$X_{f,T+1}^q = 1, Y_{f,T+1}^q = 1; \quad (6)$$

Constraints (1) reflect the requirement that the decision variables $X_{f,t}^q$ be non-decreasing in t . Thus if a flight is planned to depart from airport q by the end of time period τ , then $X_{f,t}^q$ has to be 1 for all $t > \tau$. Constraints (2) reflect the requirement that the decision variables $Y_{f,t}^q$ be non-decreasing in t . Thus if a flight is planned to arrive in airport q by the end of time period τ , then $Y_{f,t}^q$ has to be 1 for all $t > \tau$. Constraints (3) and (4) make sure that the delay time of a flight will not be affected by its pre-flight. Constraint set (5) specifies that the number of arrivals and departures during any time period is limited by the specific airport capacity for that time period. Constraints (6) ensure that all flights arrive at the destination airport or depart from the origin airport where $q \in A$ by the latest time period in the planning horizon.

III. THE NSGA-II ALGORITHM

In this section, NSGA-II algorithm [8] [9] is used to propose near-optimal solutions for the multi-objective model above. The motivation for solving the MMOM using NSGA-II algorithm is to quickly obtain a large set of potential flight-schedules which can achieve different objectives without a sharing parameter. With the properties of the fast non-dominated sorting procedure, an elitist strategy, and a parameter-less approach, the computational requirements is much lower than other algorithm, which can satisfy system requirements in practice. The overall algorithm is outlined below. We then explain each step in detail.

A. Notation

- Density Estimation:

To get an estimate of the density of solutions surrounding a particular solution in the population we take the average distance of the two points on either side of this point along each of the objectives. This quantity $i_{distance}$ serves as an estimate of the size of the largest cuboid enclosing the point i without including any other point in the population (it is called the crowding distance). In Fig. 1, the crowding distance of the i -th solution in its front (marked with solid circles) is the average side-length of the cuboid (shown with a dashed box). The following algorithm is used to calculate the crowding distance of each point in the set I .

crowding-distance-assignment(I)

$I = |I|$ number of solutions in I

for each i , set $I(i)_{distance} = 0$ initialize distance

for each objective m

$I = \text{sort}(I, m)$ sort using each objective value

$I[1]_{distance} = I[l]_{distance} = \infty$ so that boundary points are always selected
for $i=2$ to $(l-1)$ for all other points

$$I[i]_{distance} = I[i]_{distance} + (I(i+1)m - I(i-1)m)$$

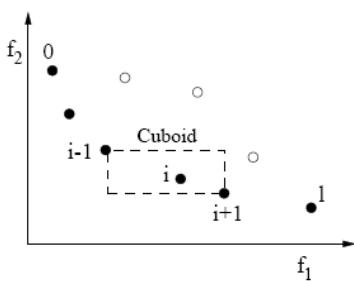


Figure 1. The crowding distance calculation is shown

- **Crowded-Comparison Operator:**

The crowded-comparison operator (\geq_n) guides the selection process at the various stages of the algorithm toward a uniformly spread-out Pareto-optimal front. Assume that every individual i in the population has two attributes:

- Non-domination rank (i_{rank});
- Local crowding distance ($i_{distance}$).

We now define a partial order \geq_n as

$$\begin{aligned} i &\geq_n j \text{ if } (i_{rank} < j_{rank}) \\ \text{or } ((i_{rank} = j_{rank}) \text{ and } (i_{distance} = j_{distance})) \end{aligned}$$

That is, between two solutions with differing non-domination ranks, we prefer the solution with the lower (better) rank. Otherwise, if both solutions belong to the same front, then we prefer the solution that is located in a lesser crowded region.

B. The Main Loop

Initially, a random parent population P_0 is created. The population is sorted based on the non-domination. Each solution is assigned a fitness equal to its non-domination level (1 is the best level). Thus, minimization of fitness is assumed. Then, tournament selection, recombination, and mutation operators are used to create a child population Q_0 of size N . From the first generation onward, the procedure is different. The elitism procedure for $t \geq 1$ and for a particular generation is shown in the following:

First, a combined population $R_t = P_t \cup Q_t$ is formed. The population R_t will be of size $2N$. Then, the population R_t is sorted according to non-domination. The new parent population P_{t+1} is formed by adding solutions from the first front till the size exceeds N . Thereafter, the solutions of the last accepted front are sorted according to \geq_n and the first N points are picked. This is how we construct the population P_{t+1} of size N . This population of size N is now used for selection, crossover and mutation to create a new population Q_{t+1} of size N . It is important to note that we use an integer tournament

selection operator but the selection criterion is now based on the niched comparison operator \geq_n .

C. Code for the NSGA-II algorithm

In order to take use of the NSGA-II algorithm in MMOM, we should code for the model first. Since the final solution is a set of feasible flight-schedules, a number of potential schedules, called a population, and the fitness of every individual in population will be introduced here.

The first step is to construct an individual. Here, that a single-level integer string represents the adjusted time period of every relative flight is used as an individual. If all flights departure from or arrival in the set of airports on time, the integer string will be a zero string. For instance, there are four flights, three of which will arrive in the opened air network while the other one will depart from it, so the individual should be connected to the arrival times of the first three flights and the departure time of the last flight. If all of them will be punctual, the individual should be 0-0-0-0. If the first flight is assigned one time period airborne delay and the last flight is assigned two time period ground delay, the individual should be 1-0-0-2. Thus, the length of the individual will be equal to the number of relative flights. Having decided on the representation, a random initialization procedure is tried for each individual to create an initial population. The size of population will be set as a parameter.

The next step is to calculate the fitness of individual. According to the individual defined above, the fitness will be decided by original schedule, objective functions and capacity constraints. The original schedule will be the basic criterion to measure the fitness. For instance, if a new individual is 1-2-0-4, so the number of adjusted flights will be 3 and the delay flights will be 1 while the delay time will be 7. Then the capacity constraints will be considered, if one of the time periods is not satisfied, three figures will be added a constant number separately as punishment. After that, with density estimation and crowded-comparison operator, the fitness will be constructed with two attributes that are non-domination rank and local crowding distance.

After coding, the NSGA-II algorithm can be used to solve the MMOM, and the flow chart is in Fig. 2.

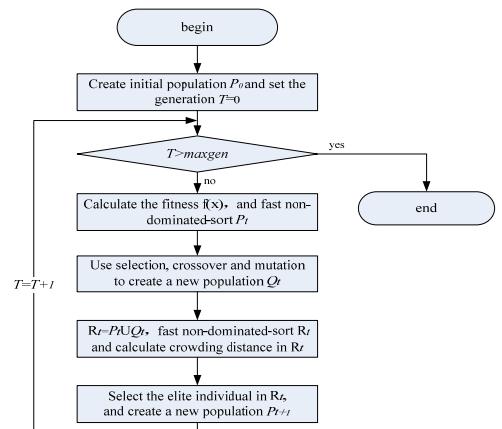


Figure 2. The algorithm flow chart is shown

IV. COMPUTATION RESULTS FOR THE NSGA-II ALGORITHM

Now we will apply the above model and algorithm to a hypothetical case involving traffic into Beijing Airport (ZBAA), Shanghai Airport (ZSSS) and Guangzhou Airport (ZGGG). For purposes of this experiment, we assumed that the three airports capacities are 60 aircraft/hour, 36 aircraft/hour and 48 aircraft/hour separately.

We consider 459 flights that are scheduled to arrive at or depart from the three airports between 8 AM and 12:05 PM on May 4, 2008. The planning horizon of 4h and 15 min is divided into 49 quarter-hour time periods. Thus $\Gamma = \{1, \dots, 49\}$ is the set of time periods, it means that every airport has 49 periods. After initial analysis, it can be seen that there are totally 35 time periods in three airports are constrained by the capacities and 57 flights are affected. In the light of First Come First Service (FCFS) basis, 176 schedules have to be adjusted, and the total time delay will be 880 minutes. Table I reports the total number of delayed flights, total number of adjusted flights, the total delay time and cost etc. In this case, we assumed that the unit airborne holding is five times as unit of ground holding. After simple calculation, the total delay cost can be seemed as the total units of ground holding.

TABLE I. THE LOST CAUSED BY FCFS

Total cost	Total delay time	Total adjusted flights	Total delay flights	Affected time periods
2260	890MIN	178	0	35

In this test problem, we use a population of size 400, a crossover probability of 0.7 and a mutation probability of 0.1. We run NSGA-II for 500 generations, and a Pareto-optimal region has been found in table II (After many times experiments, the results will not change much after generation 500 in this problem, so we use the 500 generations as the near-optimal result).

TABLE II. THE PARETO-OPTIMAL REGION IN 500 GENERATIONS

Generation	Time	Solution	Total delay cost	Total delay time	Total adjusted flights	Total delay flights
500 GENERATION	130 SECONDS	1	1735	845	73	3
		2	1705	840	62	5
		3	1650	835	67	4
		4	1790	845	70	3

We may see that, all the delay-times and delay-time-cost and adjusted flights have been cut down while the delay flights went up in order to get the near-optimal solution. Also, it can be noticed that all of the four near-optimal solutions are

more efficient than the origin schedule and also have their own advantage which can satisfy the different need from relative departments, such as the least adjusted flights need or least delayed time need.

V. CONCLUSIONS

This paper has present a new multi-objective and multi-airport optimizing model based on flights schedule. The model takes into consideration multi-airport and continued flights constraints and assigns either ground delay or airborne delays to a flight according to its departure airport. In the light of this multi-objective problem, the NSGA-II Algorithm is used to solve the model for achieving different goals with different estimated departure times.

Further research is required to address problems in which the multiple capacity constraints, including those for en-route sectors should be considered. And since this model is confined to the deterministic problem, a new optimization model which can solve the dynamic stochastic problem is the key issue of the future work.

REFERENCES

- [1] VRANAS, P. B. "The Multi-Airport Ground-Holding Problem in Air Traffic Control," Massachusetts Institute of Technology, June 1992.
- [2] Bertsimas, D., and S. Patterson, "The Air Traffic Flow Management with Enroute Capacities," Operations Research, Vol. 46, 1998.
- [3] Helme, M. P., K. S. Lindsay, S. V. Massimini, "Optimization Models for Air Traffic Flow Control Management: New Formulations," MTR 92W0000152, MITRE Corporation, 1994.
- [4] Navazio, L., G. Romanin-Jacur, "The Multiple Connections Multi-Airport Ground Holding Problem: Models and Algorithms," Transportation Science, Vol. 32, 1998.
- [5] Lindsay, K. S., E. A. Boyd, R. Burlingame, "Traffic Flow Management Modeling with the Time Assignment Model," Air Traffic Control Quarterly, Vol. 1, 1993.
- [6] Goodhart, J., "Increasing Airline Operational Control in a Constrained Air Traffic System", Ph.D. Dissertation, University of California at Berkeley, 2000.
- [7] Avijit Mukherjee, Dynamic Stochastic Optimization Models for Air Traffic Flow Management, Institute of Transportation Studies, University of California Berkeley, 2004 Paper UCB-ITS-DS-2004-3.
- [8] Omar Al Jadaan, Lakshmi Rajamani, C.R.Rao, Non-Dominated Ranked Genetic Algorithm for Solving Multi-Objective Optimization Problems: NRGA., Journal of Theoretical and Applied Information Technology, 2008 JATIT.
- [9] Kalyanmoy Deb, A Fast and Elitist Multi-objective Genetic Algorithm: NSGA-II, Transactions on Evolutionary Computation. APRIL 2002, VOL.6. No.2,
- [10] Michael O.Ball, Robert Hoffman, Amedeo Odoni, Ryan Rifkin, The Static Stochastic Ground Holding Problem with Aggregate Demands(January 1999), NEXTOR Report# RR-99-1
- [11] Sandeep Mulgund, Craig Wanke, Daniel Greenbaum, A Genetic Algorithm Approach to Probabilistic Airspace Congestion Management, AIAA Guidance, Navigation, and Control Conference and Exhibit, August 2006,Keystone,Colorado.
- [12] Michael Ball, Robert.H.Smith, Air Transportation: Irregular Operations and Control Department of Civil and Environmental Engineering, January 27, 2006.

Using Online Data to Investigate Airline Passengers' Multi-Airport Choices

Brittany Lynn Luken

School of Civil and Environmental Engineering
 Georgia Institute of Technology
 Atlanta, Georgia, United States of America
 BrittanyLuken@gatech.edu

Laurie A. Garrow, Ph.D.

School of Civil and Environmental Engineering
 Georgia Institute of Technology
 Atlanta, Georgia, United States of America
 Laurie.Garrow@ce.gatech.edu

Abstract— This paper provides my proposed research plan for my doctoral dissertation. In short, I am aiming to leverage unique internet data sources to build and validate discrete choice models in multi-airport cities. The project is divided into four steps: (1) data collection, (2) catchment area analysis, (3) discrete choice models development, and finally (4) validation of the discrete choice models, using data sources readily available to academics. From past experiences I have gained a lot of knowledge by sharing my research ideas with academics and industry professionals, and anticipate by sharing these ideas at the doctoral symposium at the ICRAT conference I will receive more valuable feedback.

Keywords- aviation; discrete choice models, air travel demand

I. INTRODUCTION

As a second year graduate student, I am finalizing a proposed plan of research for my doctoral dissertation. I am interested in investigating airline passengers' multi-airport choice in the New York region. In this paper I outline the motivation for my proposed project, how I have secured or will secure the data needed for its completion, and three distinct analyses (catchment area analysis, development of discrete choice models, and validation of discrete choice models) that will culminate to answer several pressing questions:

- Who is choosing each of the three airports in New York? When are individuals electing to go to different airports?
- What can be learned with respect to passenger preference through the development of large scale revealed preference discrete choice models spanning 20 different markets?
- Can data leveraged from a major US airline help calibrate discrete choice models developed from government data sources?

From previous experiences at other conferences, I have gained a lot of knowledge by sharing my research ideas with academics and industry professionals. I anticipate by sharing my proposed research plan during the Doctoral Symposium at

the 4th International Conference on Research in Air Transportation (ICRAT) 2010, I will receive valuable feedback.

II. MOTIVATION

According to the FAA, approximately one-third of the nation's flights and one-sixth of the world's flights either start or pass through the airspace that supports New York's three main airports: LGA, JFK, EWR. Moreover, these airports account for three-quarters of the chronic airline delays accounted for today. As noted in a recent congressional report, costs incurred by these delays are astounding: the Air Transport Association (ATA) estimates that aviation delays cost the economy \$12.5 billion a year [1]. To reduce delay, the Port Authority of New York and New Jersey has acquired a 93-year lease on Stewart International Airport, located 60 miles north of New York City, which will eventually be used to relieve passenger and cargo congestion that are overwhelming the three airports. However, without understanding of passenger preference, there is no assurance of this airport's successful implementation. According to the National Academy of Engineering, one of the grand challenges engineers of the 21st century will face is developing "integrated transportation systems, [and] making individual travel ... as easy and efficient as possible" [2]. Thus, to ensure successful implementation of the additional airport, a better understanding of how accessibility options, such as transit or roadways, influence multiple airport choice is needed. New methodologies are also critical to comprehensively solving this problem. Faced with explosive air growth on the east and west corridors of the US, the Transportation Research Board of the National Academies has stated that "high-density areas invite an entirely new approach for [air] planning and decision making that goes beyond the existing practice for transportation planning and programming" [3]. Thus, there is a need to develop new and more integrated discrete choice models that convey individuals' preferences for time-of-day departure and arrival, equipment type, cost, and landside accessibility. These models will demonstrate how to best improve accessibility to the new airport (e.g., roadway improvements or transit access) and aid in attracting appropriate airlines, such as low-cost carriers.

Outside of the policy context, this research provides an ample opportunity to calibrate models built with airline data with models build with government data sources (such as DB1A and DB1B). This will provide other academics and industry professionals with access to the DB1A and DB1B datasets and opportunity to develop more accurate discrete choice models without airline data.

III. DATA COLLECTION

Several key sources of data are needed for this analysis. Consumer purchase behavior (via ticketing records) must be captured. Choice sets available at the time of purchase are necessary to build the discrete choice models. Socio-demographic characteristics of individuals making purchases are necessary for model segmentation. Government airline data available through the Bureau of Transportation Statistics (e.g. DB1A or DB1B) is needed to validate the models. A brief description of how these sources were or will be obtained follows.

A. Ticketing Records

While working as an intern at a major US airline, I became an expert in two key software programs that I used to extract online data for a span of a year and a half for 20 different markets. This data includes a customer's IP address, frequent flyer number, zip code, loyalty status, and travel history. In terms of markets, information is available on which airports or city pairs individuals searched for flights on, days to booking, and which itineraries the individual purchased (including origin-destination pair, cabin preference, trip duration, an indicator of Saturday night stay, time of day preference, etc).

B. Available Choice Sets

During a two month period that overlaps with the year and a half long data collection at the major US airline, a simultaneous data collection was being facilitated by QL2. The QL2 data contains competitive pricing information for all major carries for the 20 US markets previously mentioned. This data enables building the choice sets necessary to develop the discrete choice models.

C. Socio-Demographic Characteristics

Georgia Tech's Center for Geographic Information Systems (GIS) has offered to share readily available socio-demographic characteristics by block group level for the northeast corridor of the United States for the completion of my dissertation. Available data includes average income, ethnicity by percentage, family structure, primary mode of transportation used to access place of employment, age distributions, and gender for each census block group.

D. Data Bank Government Airline Data

This ticketing data, commonly referred to as DB1A and DB1B, is based on a 10% sample of flown tickets collected from passengers as they board aircraft operated by US airlines. The data provides demand information on the number of passengers between origin-destination pairs, itinerary information (marketing carrier, operating carrier, class of service, etc.) and price information (quarterly fare charged by each airline for an origin-destination pair). As I get closer to using these data sources in my dissertation, I will have to contact the Bureau of Transportation Statistics to gain access to this data.

IV. CATCHMENT AREA ANALYSIS

Although catchment analysis in the New York region has previously been completed, to the best of my knowledge, it has never been broken down to the market level, or looked at over time. With the unique data set I have leveraged from a major US airline, I have developed ways to track consumers over time.

For this part of the project, I first conducted descriptive analysis of the catchment areas for the three major airports in the New York area. Through visual inspection, prominent customers to each of the three airports will be seen. Through mathematical analysis, average distances to each airport will be calculated. Specifically, the network distance will be calculated by linking zip code centroids to the roadway network provided by the Oakridge National Laboratory. Timed distance will also be calculated by using average speeds per each roadway segment. Finally, cost of travel from each zip code to each airport via the shortest path will also be factored into the analysis. Using the three abovementioned calculations (distance, time, and cost), analysis will be conducted to determine if individuals are selecting the cheapest, closest, or quickest airport to get to.

Next, individual's switching behavior will be analyzed. Specifically, factors that influence whether an individual flies out of the same airport every time he/she flies in a market will be analyzed. This will be completed by time series analysis. All markets will also be combined, so that analysis can be conducted to see if the same individuals choose different airports based on their destination choice.

Finally, linking the individuals to aggregate socio-economic data for their home zip code and other consumer purchase data, such as loyalty status or days of advance purchase, will provide opportunity for segmentation. This will enable further analysis of other driving factors for an individual's airport choice.

This part of my analysis is well under way, and I anticipate having the complete analysis done by June 2010 for the ICRAT conference.

V. DEVELOPMENT OF DISCRETE CHOICE MODELS

Many multi-airport choice models have been developed for cities across the United States in recent years. Some have used stated preference survey data while others have compiled small sets of disaggregate data. This research is unique in that it leverages various distinct sources of data (such as disaggregated revealed preference data for a US airline's passengers as well as competitor's flight level pricing data).

To complete this analysis, the alternative choice sets will have to be built. Since the ticketing data indicates days of travel, and day of purchase, correct choice sets can be developed for each individual based on the QL2 pricing data. Although this will be the most complicated part of the analysis, I am confident that computer programs can be written that build every choice set. Subsequently, each choice set will then need to be linked back to the ticketing data. With the linked data, discrete choice models can be run using Easy Logit Modeling (ELM) or Statistical Analysis Software (SAS) fairly easily.

This research looks to pave the way for analyzing an unprecedented large set of disaggregate data. It will be distinctive since it will provide the first revealed preference multi-airport itinerary choice model based on detailed pricing data. I hypothesize that the unique online search and purchase data representing millions of individual consumer purchase transactions will more significantly reveal individual's willingness to pay for different attributes (e.g. airport choice, time of departure, accessibility, etc.) associated with the product. Thus, this research will indicate how influential accessibility options are to an individual's airport choice and in turn indicate where investments should be made with respect to landside accessibility to existing and new airports in a multi-airport region.

VI. VALIDATION OF DISCRETE CHOICE MODELS

With the development of discrete choice models using an unprecedented culmination of data sources, comes the responsibility to help others in the field replicate the research using broadly available data sources. Thus, after development of discrete choice models (as indicated in Part V), I hope to redevelop the same models using government data sources (DB1A and DB1B). I then want to calibrate the aggregate models developed with government data sources to the disaggregate models created with the data leveraged from a US airline. Although, I am uncertain how this calibration will be completed, I believe this investigation will be the most valuable to the research community once completed.

VII. CONCLUSION

I approach my dissertation project with both enthusiasm and caution. I am excited to see how extracting large volumes of disaggregate internet data can improve a modeler's ability to predict an individual's preferences. I am also eager to develop ways to calibrate my models to models that other researchers could repeat using government data sources. However, I am cautious because in all research unexpected hurdles are presented. I hope through sharing my ideas with fellow doctoral students, academic professors and industry professionals, I will be able to dismantle some of the hurdles early in my research process. I also hope that through the symposium I will further develop some of the research ideas presented in this paper, and help others in the community further their own.

VIII. BIOSKETCH

Brittany Lynn Luken is a doctoral student at the Georgia Institute of Technology. She is pursuing a Ph.D. in Civil Engineering, specializing in Transportation Engineering with her research focus on travel demand modeling in aviation. Brittany graduated cum laude from Vanderbilt University in May of 2008. She has recently been selected as one of ten recipients nationally of the Graduate Research Award Program for Airport Cooperative Research Group for the academic year 2009-2010. Brittany is also recipients of a Georgia Department of Transportation (GDOT) Scholarship, scholarship to attend the AV020 mid-year meeting, and the Gordon Schultz Scholarship.

ACKNOWLEDGMENT

Funding for the catchment area analysis (Part IV) has been provided by an Airport Cooperative Research Program Graduate Research Fellowship I received for the 2009-2010 academic year.

REFERENCES

- [1] Congestion Management in New York Air Space. Transcript to the Members of the Subcommittee on Aviation, US House of Representatives. 17 June 2008.
- [2] National Academy of Engineering. 2009. Grand Challenges for Engineering.
- [3] ACRP 03-10: Innovation Approaches to Addressing Aviation Capacity Issues in Coastal Mega-Regions. 2008.

A Study of Characteristics of Solutions Obtained by Heuristics for Regional Air Traffic Flow Management

Ying Zhang

College of Civil Aviation

Nanjing University of Aeronautics and Astronautics

Nan Jing, China

yoyozhying@163.com

Minghua Hu

College of Civil Aviation

Nanjing University of Aeronautics and Astronautics

Nan Jing, China

minghuahu@263.net

Abstract—In order to accommodate the growing demands for regional air traffic flow management decision support, this paper firstly presents an integer programming model of the problem. For the purpose of solving the problem in real time, heuristics is used to reduce the computational complexity. This heuristic method divides the capacity restriction problem of multiple-unit into a sequence of capacity restriction problems of single-unit. The algorithm will converge. It is found that the locality where the flight delay happens will have an impact on the result of the algorithms, which is measured as the total delay for all the flights. Two different delay strategies are compared in different sector capacity conditions. It is found that either single delay strategy is always more efficient than the other under any airspace conditions. Therefore, it is necessary to add heuristic information concerning delay distribution to improve the heuristics applied in this paper.

Keywords-regional air traffic flow management; heuristics; integer programming

I. INTRODUCTION

The past decade has witnessed the rapid development of China's air transportation industry. For example, in 2006, the total air traffic volume of China ranks 2 in the world. However, the air traffic flow is distributed unbalanced in the whole country. Almost 70% of the traffic flow is distributed in the eastern triangle area with three vertexes, BJ, SH, and GZ. The congestion has become a problem for the reasons of economy and safety. Three area control centers have been built in Beijing, Shanghai and Guangzhou.

Air traffic flow management (ATFM) plays an important role balancing the air traffic flow with the capacity. Regional traffic flow management (RTFM) plan is local-oriented and more detailed. It is planned on a forecast time horizon of roughly 20 minutes to 2 hours and with more accurate traffic and weather information. The two primary flow control strategies for RTFM are miles-in-trail (MIT) restrictions and local reroutes. The objective of the RTFM is to find a feasible solution to minimize a particular cost function.

Currently in the three area control centers, traffic flows are controlled through Miles-in-Trail (MIT) or Minutes-in-trail (MINIT) in order to prevent congestion in en-route sectors and arrival airports. Flow restrictions are passed back from passed from the downstream airspace unit to the upstream airspace unit as needed. In present-day operations, because the restrictions are imposed by the TFM managers, and they are

always restricted by the experience of the staff, the restrictions are often found passive and cause new congestions. Pre-tactical decision support tools for the RTFM problem are urgently needed.

II. RELATED WORK

The problem of decision-making in ATFM can be solved with linear or non-linear optimization techniques. There exists a moderate amount of research literature on applying optimization methods to TFM decisions. The optimization model in [1] is most computationally efficient compared to other models, but it becomes computationally intractable for large-scale TFM problems. A deterministic optimization model which can be seen as a "macroscopic version" of the model presented in [1] is presented in [2] for the European ATFM problem. The model illustrates the complex nature of European ATFM solutions. Optimal departure delays of the flights obtained from the deterministic optimization model is used in [3] to back-calculate the necessary MIT restrictions at the Center boundaries which can be used for the regional TFM.

Since optimization methods sometimes can't meet the computation time requirement, heuristic methods can be applied in order to obtain suboptimal solutions reasonably fast. Four algorithms including two heuristic methods, an optimization method and a hybrid method, are formed and evaluated in [4].

III. MODELING AND ALGORITHMS FOR REGIONAL TFM PROBLEM

A. Problem modeling

1) The assumptions and notation of the model

The model is based on the following assumptions:

- The arrival and departure capacity of a single airport is restricted. Only the airport acceptance rate and departure rate of the busiest airports in the region are considered.
- No consideration is given to delay propagation from earlier flights to subsequent flights by the same flight.
- The maximum amount of delay that can be assigned to any flight is not restricted.
- The delay occurs before the flight enters the target region or when the flight is in the region.

The notations are defined as follows:

T: the set of time intervals

K: the busiest airport in the region

F: the set of flights

$F_a^K \subset F$: the set of flights whose destination airport is K

$F_d^K \subset F$: the set of flights whose departure airport is K

d_f : the scheduled departure time for flight f

a_f : the scheduled arrival time for flight f

S: the set of sectors in the target region

P(f,i): the ith sector that flight f will transit, $1 \leq i \leq N_f$

$l_{fp(f,i)}$: the minimum time for flight f to cross sector P(f,i)

AAR(t): the airport acceptance rate in the time period t

ADR(t): the airport departure rate in the time period t

δ : time slot for the sector capacity constraint (minutes)

$PE^s = \{pe_0^s, pe_1^s, \dots\}$: successive periods of length δ for

sector s, each with a start and an end, $start(pe_l^s) - end(pe_l^s) = \delta$

$C_{pe_l^s}$: the period capacity of pe_l^s for sector s, $s \in S$

w_{fj} : the decision variables for flights $f \in F$ where

$$w_{fj} = \begin{cases} 1 & \text{if flight } f \text{ arrives at sector } j \text{ by time } t \\ 0 & \text{otherwise} \end{cases}$$

y_{ft} : the decision variables for flights $f \in F_a^K$ where

$$y_{ft} = \begin{cases} 1 & \text{if flight } f \text{ lands at time } t \\ 0 & \text{otherwise} \end{cases}$$

z_{ft} : the decision variables for flights $f \in F_d^K$ where

$$z_{ft} = \begin{cases} 1 & \text{if flight } f \text{ takes off at time } t \\ 0 & \text{otherwise} \end{cases}$$

G_f : the ground delay for flight f

$$G_f = \begin{cases} \sum_t z_{ft} t - d_f & f \in F_d^K \\ 0 & f \in F / F_d^K \end{cases}$$

A_f : the air delay for flight f

$$A_f = \begin{cases} \sum_t y_{ft} t - a_f - G_f & f \in F_d^K \\ \sum_t y_{ft} t - a_f & f \in F / F_d^K \end{cases}$$

$C_f^g(G_f)$: the ground delay cost function for flight f, it can be linear or non-linear functions of the ground delay, here it's thought of as the sum of a geometric progression for each increase of the ground delay, α is the geometric proportion, b_f is the cost of the first unit ground delay^[5].

$$C_f^g(G_f) = \sum_{k=1}^{G_f} (b_f (1 + \alpha)^{k-1}) \quad \alpha \neq 0$$

$C_f^a(G_f, A_f)$: the air delay cost function for flight f, it can be linear or non-linear functions of the air delay, it is the sum of a geometric progression for each increase of the air delay, k is the ratio of air delay cost to ground delay cost^[5].

$$C_f^a(G_f, A_f) = \sum_{k=1}^{A_f} (k * b_f (1 + \alpha)^{G_f + k - 1})$$

2) Constraints and Optimization Criterion:

$$\text{Min} \sum_{f \in F} [c_f^g(G_f) + c_f^a(A_f)]$$

Subject to:

$$\sum_{t \in T} y_{ft} = 1 \quad \forall f \in F_a^K \quad (1)$$

$$\sum_{t \in T} z_{ft} = 1 \quad \forall f \in F_d^K \quad (2)$$

$$\sum_{f \in F_a^K} y_{ft} \leq AAR(t) \quad \forall t \in T \quad (3)$$

$$\sum_{f \in F_d^K} z_{ft} \leq ADR(t) \quad \forall t \in T \quad (4)$$

$$\sum_{t=start(pe_l^j)}^{end(pe_l^j)-1} \sum_f (w_{fij} - w_{f,t-1,j}) \leq C_{pe_l^j} \quad (5)$$

$$w_{f,t+l_{fj},j'} - w_{fij} \leq 0, j = P(f, i), j' = P(f, i+1) \quad (6)$$

$$w_{fij} - w_{f,t-1,j} \geq 0 \quad (7)$$

The model is expressed as an integer programming model. The objective function is the sum of the ground and air delay cost. Constraints (1) and (2) are the assignment constraints that uniquely determine the arrival time for each flight $f \in F_a^K$ and the departure time for each flight $f \in F_d^K$. Constraints (3) and (4) are the airport arrival rate and airport departure rate constraints. Constraint (5) is the sector period capacity constraints which are defined as the maximum flight numbers that can enter the sector during the period. Constraint (6) represents connectivity between sectors. Constraint (7) represents connectivity in time. The proposed model is based on the Bertsimas and Patterson's (1998), however the sector capacity here is expressed as period capacity instead of the instantaneous capacity and the targeted area of application is more local in space. The period capacity will not be too restrictive and it has different forms such as non-overlapping window period or sliding period to get different effects^[6].

B. Algorithms for the model

Heuristic method is used to solve the problem for its simplicity and short runtime. The idea is motivated by [3] whereas the airport capacity restriction is added. Firstly the capacity restriction problem of multiple-unit is divided into a sequence of capacity restriction problems of single-unit to be solved iteratively. The sequence of the airspace unit is arranged according to the congestion degree of the unit. The airport congestion is solved before the sector. For the capacity restriction problem of single-unit, delay is assigned to the flights which will cross the unit to meet the capacity restriction and minimize the delay cost. Each time the single-unit problem is solved, traffic demands are recomputed to take into account the delay assigned. Then the sequence of the airspace unit is reordered. The most congested airspace unit is chosen, and then repeat the process described above. The iteration ends when all the units are within their capacity and the algorithm will converge. For each single-unit problem, the First-

Scheduled-First-Served (FSFS) rule which is thought of as a fairness criterion in ground delay program (GDP) is used.

IV. RESULTS

Heuristics is applied to the Central South Area air traffic to explore characteristics of solutions obtained by the algorithm. Although it doesn't matter where the delay happens for the single-unit problem, it does count for a lot for the multiple-unit problem. The result, which is measured as the total delay for the flights, is examined for the following two delay distribution strategies:

Strategy 1: Delay happens before flight enters the whole region or on the ground for flight which departs from the airports in the region.

Strategy 2: Delay happens in the region as well as outside of the region. Delay propagates from the close upstream airspace unit of the congested unit to the further upstream airspace unit.

In the Central South Region, there contains 16 area-control sectors (see Figure 1) and 5 terminal-control sectors including AN, AE, AW, ARR, and DEP. These sectors handle traffic to and from some major airports:

- Flights which will only transit the area sectors;
- Arrival to / Departure from Guangzhou Baiyun Airport (ZGGG) which will pass both the area sectors and the terminal sectors. For example, flights arrive at ZGGG will pass the area-control center and two of the terminal-control sectors which include AN or AE or AW and ARR;
- Arrival to / Departure from Baoan (ZGSZ), Zhuhai/Sanzao (ZGSD) and Macau (VMMC) which will pass both the area sectors and the terminal sectors including AE or AW or AN.

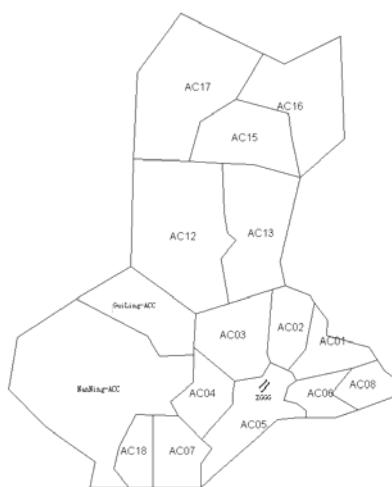


Figure 1. The area control sectors of Guangzhou Area. The busiest airport in the area is Guangzhou Baiyun International Airport (ZGGG).

In order to compare the two strategies, the following two conditions are examined and the flights which only arrive at ZGGG are considered to simplify the comparison:

- Condition 1: The adjacent sectors the flights are going to transit are all congested;
- Condition 2: There are uncrowded sectors between congested sectors the flights are going to transit.

To realize condition 1, the two capacity restricted sectors are AN and ARR which are adjacent sectors. To realize condition 2, the two capacity restricted sectors are AC02 and ARR while AN is the capacity unrestricted sectors between AC02 and ARR. For both of the two conditions, five groups of different capacity value are tested for the two delay strategies (see Table I and Table II). Both the capacity value (flights/15 minutes) of the sectors and the total delays (minutes) are listed in the table.

TABLE I. TOTAL DELAY OF THE TWO DELAY STRATEGIES UNDER DIFFERENT CAPACITIES IN CONDITION 1

AN	ARR	Strategy 1	Strategy 2
5	5	1569	1810
6	6	1055	1103
5	8	617	654
8	5	1569	1569
8	8	405	405

TABLE II. TOTAL DELAY OF THE TWO DELAY STRATEGIES UNDER DIFFERENT CAPACITIES IN CONDITION 2

AC02	ARR	Strategy 1	Strategy 2
8	8	405	405
5	8	506	488
6	6	1055	1055
5	5	1582	1569
3	5	1998	1862

In condition 1, since the delay value in Strategy 1 is less than or equal to the delay value in Strategy 2, Strategy 1 is more efficient than Strategy 2. Chances are that flight delays in the upstream will just increase the crossing time of the flight and make the upstream sectors more congested. In condition 2, Strategy 2 is more efficient than Strategy 1. The explanation is that the uncrowded airspace unit can be used as buffer area for the flight to use before reaching the congested airspace unit. So either delay strategy is always more efficient than the other.

V. CONCLUSIONS AND FUTURE WORK

The model for the RTFM focuses on minimizing the total delay cost while balancing the airspace sector and airport capacity in the region with the air traffic flow. Heuristic method is chosen for solving the problem because of its intuitive simplicity, short runtime, and compatibility with the FSFS principles. Although the results obtained by the method are sub-optimal, it can be used as a basis for comparison. And it can also be used as an initial feasible solution for the integral model.

Since it is found delay strategy have impact on the results, the heuristics can be improved by adding heuristic information concerning delay distribution. This is a direction for future research.

REFERENCES

- [1] Bertsimas, D., and Patterson, S.S., "The Air Traffic Flow Management Problem with En Route Capacities", Operations Research, Vol. 46, 1998, pp. 406-422.
- [2] Lulli, G., Odoni, A.R., "The European Air Traffic Flow Management Problem", Transportation Science 41,1-13(2007)
- [3] A Mukherjee, S Grabbe, Banavar Sridhar, "Alleviating Airspace Restrictions through Strategic Control", AIAA Guidance, Navigation and Control Conference and Exhibit, 18-21 August 2008, Honolulu, Hawaii
- [4] Jakobovits ,R., Kopardekar, P., Burke, J., and Hoffman,R., "Algorithms For Managing Sector Congestion Using The Airspace Restriction Planner", 7th USA/Europe ATM R&D Seminar, Barcelona, Spain, July 2007
- [5] Minghua Hu, Xiaohao Xu,Aiming Cheng, "Multiple unit ground holding strategy problem research in air traffic flow management ", Aeronautica Et Astronautica Sinica, Vol. 19 No. 1, Jan. 1998
- [6] Nicolas Barnier, Pascal Brisset and Thomas Riviere., "Slot allocation with constraint programming: models and results", Proceedings of the fourth USA/Europe Air Traffic Management R&D Seminar, 2001

A Mixed Integer Linear Model for Potential Conflict Minimization by Speed Modulations

David Rey, Sophie Constans, Rémy Fondacci

Université de Lyon F-69622, Lyon, France

LICIT, INRETS, ENTPE F-69675 Bron, France

Email: {david.rey, sophie.constans, remy.fondacci}@inrets.fr

Christophe Rapine

Laboratory G-SCOP

46 avenue Felix Viallet, Grenoble F-38031, France

Email: christophe.rapine@g-scop.fr

Abstract—As air traffic volume is growing around 5% each year in Europe it has become a priority to improve air traffic control in order to deal with tomorrow's air traffic configuration. In 2007 the SESAR (Single European Sky ATM Research) project was created under European Community law as an initiative to design the future of air traffic management over Europe. One of the objective of SESAR is to increase air traffic density and optimize flight route plans. This can be achieved through en-route deconfliction. Reducing the global number of conflicts through speed regulation has been tested in the ERASMUS project (En-Route Air traffic Soft Management Ultimate System) and efficiency of the concept has been analysed through tests and simulations [2]. It provided many insights related to en-route control using speed variations hence becoming a solid reference. In this paper we develop a mixed integer linear model for a speed regulation problem that suits SESAR requirements. We focus on flights crossing times at intersection points rather than distance to ensure separation along our resolution. Speed regulation is thus converted into travel time control. Finally we propose an integer linear program aiming at minimizing the global potential conflict quantity by speed modulations.

Index Terms—ATM, Conflits, Mixed Integer Programming, Speed Control.

I. INTRODUCTION

ATM (Air Traffic Management) has become a challenge in Europe since the beginning of the 90's when it started to increase over 5% yearly. Airspace capacity is extended year after year in order to deal with growing demand, slowly reaching its limits. Europe's core area has one of the highest air traffic density in the world and generates many potential air conflicts every day. An air conflict occurs when two aircraft are too close to each other resulting in a loss of separation, which is a dangerous situation. Managing aircraft separation requires a high cognitive activity and directly affects ATCos' (Air Traffic Controllers) capacity to deal with large numbers of flights, therefore reducing the global airspace capacity. Hence today's ATM research and development is looking forward to reduce ATCos' workload by developing new tools to help controllers managing greater quantities of flights simultaneously while ensuring a high safety level. The ERASMUS project was funded to develop efficient ways to integrate advanced automation concepts in the SESAR framework. One of the objective of the SESAR Master Plan is to optimize flight route plans and as suggested in [2], automatic speed control could

be a way to improve air traffic deconfliction.

Many conflict detection and resolution methods have been tested and reviewed in [5]. Among the conflict resolution methods, lateral and vertical maneuvers are described in [3] and in [4]. We focus on minimizing the global conflict quantity through speed regulation. We chose a *global* multiple aircraft conflict resolution method as described in [5] thus considering the whole traffic simultaneously, using a mixed integer linear model. The conflict detection part is done using flight intentions and existing routes. Merging all this data, it is possible to come up with a map containing all intersections of all flight trajectories and thus all potential conflicts. Conflict resolution through speed regulation is currently not used in practice by ATCos. Indeed, ATCos rarely suggest speed variation to solve a potential conflict as it is generally not appreciated by pilots and airliners for fuel consumption and delay matters. Similarly, it is very difficult for an ATCo to perceive speed regulation on a radar screen making the automation of speed control a natural solution. In the next section we describe the framework of our resolution, then we introduce our model for the speed regulation problem and conclude with some perspectives.

II. RESOLUTION FRAMEWORK

The method developed here was first introduced in [1], we recall part of the original setup for it sets the basics of our model.

To reach their destination, aircraft follow trajectories that frequently intersect and overlap each other, especially in dense areas. Most conflicts happen around such intersection points thus we choose to focus on all those geographical references. The first step in our resolution method is to locate all intersection points of all routes on a two-dimensional map. Above each of these intersection points we identify the altitudes of all flights passing near the current intersection point in order to obtain a three-dimensional map of all potential conflict zones. If two aircraft are below the vertical separation norm in the neighborhood of an intersection point, they are considered in a potential conflict situation. We now define the notion of *conflict cost*. This notion is to be understood from the ATCo's point of view. Indeed, the longer two aircraft are predicted to be in conflict, the greater the monitoring and conflict solving workload is. Therefore,

the more time a controller spends to monitor and solve a conflict, the more expensive it should be. Defining a conflict cost in accordance with ATCos' workload seems to be the right manner to characterize the severity of a conflict. From a mathematical point of view, we define the cost of a conflict as a function of the time two aircraft spend under the separation norm. Adding up the costs of all potential conflicts over a given region we obtain a global conflict indicator for this region. Our objective is to minimize this indicator using speed modulations on concerned aircraft.

Flight trajectories, intersection points and aircraft performances are used as input data in our model. After calculating the theoretical arrival times of all flights on all intersection points and all the travel times from an intersection point to the next one, the next step is to optimize those arrival times through speed modulations in order to minimize the global conflict indicator. Recalling the method used in [1], we modify travel times to express speed variations. Indeed, bounding the speed variation range is equivalent to bounding the travel time from an intersection point to another for a given aircraft. Due to aircraft engine characteristics, the optimal speed variation range for an Airbus A320 is -6% to $+3\%$ of the nominal en-route speed according to [2]. Although this is the recommended speed variation range, it should be considered as one of the optimization problem parameter as well as the separation norm. As described in [1], we suppose that a sliding horizon loop with a fixed step is used to continuously regulate the traffic flow. The regulation time step has to be chosen carefully since the optimization problem is solved during each iteration. The optimization horizon (*i.e.* the period of time affected by the optimization) is to be determined together with the regulation step to limit the overall optimization problem size as it may reach extremely long computation times. Simulations are thus required to achieve better tuning of these parameters. The output data of our optimization problem are the optimized arrival times of all flights at intersection points in the optimization horizon. In the next section, we introduce our mathematical formulation for the speed regulation problem.

III. MATHEMATICAL FORMULATION

As mentioned in the previous section, the cost of a conflict depends on the configuration of the intersection as well as the aircraft speeds. In the en-route airspace the conflict area is a cylinder of 5 NM radius and 1000 ft height centered on the aircraft. Any other aircraft entering this area is considered “in conflict” with the current aircraft. In order to be more compatible with existing and future technology in ATM, we define the notion of conflict using a time interval. Focusing on the difference of the crossing times at a given intersection point of two aircraft, we can ensure separation. Quoting [1], we say that flights f and f' are in conflict at intersection point i if the instants t_{fi} and $t_{f'i}$ when flights f and f' cross

intersection point i , are such that:

$$|t_{fi} - t_{f'i}| < \sqrt{\frac{D_s^2(u^2 + u'^2 + 2\alpha uu')}{u^2 u'^2 (1 - \alpha)^2}} \quad (1)$$

where u and u' are the flights speeds, α is the cosinus of the angle between their trajectories and D_s the separation norm (*i.e.* 5 NM) parameter. The case where both aircraft share a common segment (*i.e.* $\alpha = 1$) is not described here. Information on this case can be found in [1].

Let $\Gamma_{ff'i}$ be the right hand side of (1). We define the cost of a potential conflict between flights f and f' at intersection point i such as:

$$\max(\Gamma_{ff'i} - |t_{fi} - t_{f'i}|, 0) \quad (2)$$

This local cost vanishes as soon as $|t_{fi} - t_{f'i}|$ is large enough to ensure separation between flights f and f' , and the closer the crossing times are, the higher the cost is. Defining the global cost as the sum of all the local ones, for all the flights and all the intersection points, our objective is to minimize this cost. In order to achieve en-route speed regulation on aircraft flying on intersecting routes, we focus on travel times from one intersection point to the next one. Let τ_{fi} be the travel time of flight f from its $(i-1)^{th}$ to i^{th} intersection point, we get the relations (3) and (4):

$$t_{fi} = t_{fi-1} + \tau_{fi}, \quad \forall i \neq 1, \quad i \in I_f \quad (3)$$

$$t_{f1} = T_{fREF} + \tau_{f1} \quad (4)$$

Where I_f is the set of all intersection points of flight f . T_{fREF} is equal to T_{REF} , the time when the picture of the traffic is taken (*i.e.* the reference instant), if the aircraft is already airborne, and to the flight take-off time of flight f otherwise. Equations (3) and (4) can be rewritten in a matricial form:

$$Bt_f = \tau_f + T_{fR} \quad (5)$$

where T_{fR} is a vector of zeros, except for its first component, equal to T_{fREF} , and where B is a square matrix such that $b_{ij} = 1$ if $i = j$, $b_{ij} = -1$ if $i-1 = j$ ($i > 1$), and $b_{ij} = 0$ otherwise.

Since we want to control the aircraft speeds along their trajectories, we have to bound the aircraft speed variation range. This is done by bounding their travel times. Let τ_f^m and τ_f^M be two vectors to limit the range of these modifications for flight f as lower and upper bounds on τ_f . We are now able to reproduce the optimization problem developed in [1]:

$$\begin{aligned} \min \sum_{i \in I} \left(\sum_{\substack{f, f' \in F_i \\ f < f'}} \max \left(\Gamma_{ff'i} - |t_{fi} - t_{f'i}|, 0 \right) \right) \\ \text{s.t. } \tau_f^m + T_{fR} \leq Bt_f \leq \tau_f^M + T_{fR}, \quad \forall f \in F \end{aligned} \quad (6)$$

where $t_{fi}, t_{f'i} \in \mathbb{R}^+$ are the decision variables of the problem, I is the set of the intersection points, F_i the set of the flights crossing intersection point i and F the set of all flights.

This formulation is not linear because of the absolute value in the objective function. In order to linearize this expression we introduce new parameters. Let t_{fi} and \bar{t}_{fi} be the minimal and maximal instants possible for flight f to cross intersection point i according to the speed variation range. We define two binary variables $y_{ff'i}$ and $y_{f'fi}$ to characterize the crossing order between flights f and f' at intersection point i . Variable $y_{ff'i}$ is such as:

$$y_{ff'i} \equiv \begin{cases} 1 & \text{if } t_{fi} \leq t_{f'i} \\ 0 & \text{otherwise} \end{cases}$$

and reciprocally for $y_{f'fi}$.

Consider the following optimization problem with decision variables $\omega_{ff'i}$, t_{fi} , $t_{f'i}$, $y_{ff'i}$ and $y_{f'fi}$:

$$\min \sum_{\substack{f, f' \in F_i \\ i \in I \\ f < f'}} \omega_{ff'i} \quad (7)$$

subject to:

$$\tau_f^m + T_{fR} \leq Bt_f \leq \tau_f^M + T_{fR}, \quad \forall f \in F \quad (8a)$$

$$\omega_{ff'i} \geq \Gamma_{ff'i} - (t_{fi} - t_{f'i}) - (\bar{\Gamma}_{ff'i} + (\bar{t}_{f'i} - \underline{t}_{fi})) \cdot y_{ff'i} \quad (8b)$$

$$\omega_{ff'i} \geq \Gamma_{ff'i} - (t_{f'i} - t_{fi}) - (\bar{\Gamma}_{ff'i} + (\bar{t}_{fi} - \underline{t}_{f'i})) \cdot y_{f'fi} \quad (8c)$$

$$t_{f'i} \leq t_{fi} + (\bar{t}_{f'i} - \underline{t}_{fi}) \cdot y_{ff'i} \quad (8d)$$

$$t_{fi} \leq t_{f'i} + (\bar{t}_{fi} - \underline{t}_{f'i}) \cdot y_{f'fi} \quad (8e)$$

$$1 = y_{ff'i} + y_{f'fi} \quad (8f)$$

$$\omega_{ff'i}, t_{fi}, t_{f'i} \in \mathbb{R}^+ \quad (8g)$$

$$y_{ff'i}, y_{f'fi} \in \{0, 1\}. \quad (8h)$$

We show that this mixed integer program is a correct formulation of optimization problem (6). Consider for instance the case where flight f crosses intersection point i before flight f' : we have $t_{fi} \leq t_{f'i}$ thus $y_{ff'i} = 1$. Constraint (8b) becomes:

$$\omega_{ff'i} \geq (t_{f'i} - \bar{t}_{f'i}) + (\underline{t}_{fi} - t_{fi})$$

which right-hand side is negative and thus vanishes because of (8g) (by redundancy). Constraint (8c) becomes:

$$\omega_{ff'i} \geq \Gamma_{ff'i} - (t_{f'i} - t_{fi})$$

Since $\omega_{ff'i}$ only appears in constraints (8b), (8c) and (8g), its value at the optimum is:

$$\omega_{ff'i}^* = \max(\Gamma_{ff'i} - (t_{f'i} - t_{fi}), 0).$$

This last expression of $\omega_{ff'i}$ is equivalent to the main part of the objective function in optimization problem (6). Similarly we can obtain the transposed formulation if $t_{f'i} \leq t_{fi}$. Therefore, optimization problem (7) subject to the set of constraints (8a) to (8h) is a linear formulation of optimization problem (6). This mixed integer program can now be implemented on a linear programming solver such as CPLEX.

IV. CONCLUSION

We developed an air conflict resolution method using a mixed integer linear model designed to solve the speed regulation problem in air traffic control. Experimentations are now required to set the limits of our approach, in particular to determine the maximum flight density we can deal with. As many mixed integer problems can have exponential computation times, we should watch carefully the program total running time when implemented on real-size scenarios, in order to cope with SESAR requirements. Tuning of the optimization problem parameters is essential to better achieve deconfliction through speed modulations, hence tests and simulations are to be performed in a near future. Results from these experimentations should guide us in our future work and could be compared to other results achieved in the ERASMUS project.

REFERENCES

- [1] Sophie Constans, Bastian Fontaine, and Rémy Fondacci. Minimizing potential conflicts with speed control. In *International Conference on Research in Air Traffic Transportation, IC RAT*, 2006.
- [2] Fabrice Drogoul, Philippe Averty, and Rosa Weber. ERASMUS strategic deconfliction to benefit SESAR. In *Eighth USA/Europe Air Traffic Management Research and Development Seminar*, 2009.
- [3] Geraud Granger, Nicolas Durand, and Jean-Marc Alliot. Optimal resolution of en route conflicts. In *ATM 2001*, 2001.
- [4] Richard Irvine. Comparison of pair-wise priority-based resolution schemes through fast-time simulation. In *Eighth Innovative Research Workshop & Exhibition, EEC*, 2009.
- [5] James K. Kuchar and Lee C. Yang. A review of conflict detection and resolution modeling method. Technical report, Massachusetts Institute of Technology, 2000.

Using On-Line Data to Explore Competitive Airline Pricing Policies

Stacey Mumbower and Laurie A. Garrow

Georgia Institute of Technology
 School of Civil and Environmental Engineering
 Atlanta, Georgia, United States
 stacey.mumbower@gatech.edu

Abstract— Since the mid 2000's, the airline industry has seen volatile fuel prices, a record number of carriers ending service, and a merger between two major airlines. In a time of such turmoil in the industry it is increasingly important to understand the relationship between airline consolidation and competitive pricing policies, as this relationship will directly impact the formation of future airline policies associated with competition policy (anti-trust), deregulation, and mergers. However, there is a lack of consensus about market concentration and its influence on airfares, mainly due to data limitations of past research. Given the emergence of on-line booking engines, there is a new opportunity to collect detailed fare data. This project uses disaggregate, on-line airfare data to study the relationship between market concentration and pricing policies. The dataset includes 62 markets that cover a broad range of market structures. A case study approach is used to analyze the data. Using disaggregate fare data, this study finds low price dispersion can be associated with both low and high levels of market concentration. As the day of departure approaches, price dispersion is seen to either increase or decrease, depending on the specific market. Additionally, peak and off-peak periods demonstrate differing pricing strategies. Also, markets with codeshares are shown to sometimes exhibit unusually high price dispersion.

Keywords-price dispersion, competitive pricing, market structure

I. INTRODUCTION

The research presented in this paper was completed during Mumbower's first year in the Ph.D. program at Georgia Tech and serves as a foundation for her dissertation. In presenting this research at the Doctoral Symposium of the 2010 International Conference on Research in Air Transportation, the authors anticipate that the feedback from the research community will help guide the future of this research in a direction that would be most valuable to the research community.

II. MOTIVATION

Since its deregulation in 1978, the airline industry has seen a large number of changes. Low cost carriers (LCCs) have penetrated the market and generally offer lower prices than legacy carriers. The internet has transformed the industry by greatly increasing pricing transparency. On-line travel agents

such as Expedia®, Orbitz®, and Travelocity® make it convenient for customers to search the prices of multiple airlines across multiple departure dates. Customers can find and purchase the lowest possible fare in a matter of minutes. The growth of LCCs combined with the increased transparency of airfares has led, at least in part, to lower average prices in the airline industry. Between 1995 and 2004, the prices that passengers paid for tickets declined by more than 20 percent after adjusting for inflation (5). While decreased prices are good for consumers, its implications on airlines are quite the opposite. Airline operating costs have increased dramatically over the last few years, but airlines have not been able to increase fares to match rising costs. In the first quarter of 2009, U.S. network carriers reported a total operating loss of \$867 million, which was the sixth consecutive quarterly loss (6). Between 2002 and 2008, four major U.S. carriers filed for bankruptcy protection (Delta Air Lines, Northwest Airlines, United Airlines, and US Airways). In addition, ATA Airlines, Skybus Airlines, and Aloha Airlines filed for bankruptcy and ended service. Frontier Airlines has also filed for bankruptcy but has not ended service, and in 2008 Delta and Northwest merged in an effort to become more financially stable.

In a time of such turmoil in the industry it is increasingly important to understand the relationship between airline consolidation and competitive pricing policies, as this relationship will directly impact the formation of future airline policies associated with competition policy (anti-trust), deregulation, and mergers. In the past, many researchers have studied how market structure affects the dispersion of airfares, often called price dispersion. Price dispersion has been defined in many ways by different researchers and is specific to the unit of observation of analysis. However, price dispersion can generally be thought of as the difference between an airline's highest and lowest fares in a market.

III. LITERATURE REVIEW

The interest in price dispersion of airfares was sparked in 1989 when Borenstein used government data sources to show that there is a negative relationship between market concentration and price dispersion (7), meaning that as a route becomes more dominated by one airline and moves closer

towards monopoly the price dispersion decreases. More specifically, he found that as a route moves closer towards a monopoly, an air carrier's low-end fares increase while high-end fares decrease, thus decreasing the overall dispersion of prices (while increasing average prices). Over the next several years, other researchers also used U.S. government data sources to study this relationship empirically, with findings that supported the negative relationship between market concentration and price dispersion (8, 9, 10). A theoretical model also supported this relationship (11). These researchers also found many other factors that influence the dispersion of prices. For instance, it has been shown that price dispersion increases with increased airport dominance (8), airport congestion (8), and internet search for airfares (10). Price dispersion has also been shown to decrease with increased frequency of flights on a route (8), higher levels of tourist traffic (8), and competition from Southwest (9).

The negative relationship between market concentration and price dispersion has been contradicted, however, in at least two more recent studies that use the same government data sources and analyze the data differently. In past studies, the modeling approach was to take millions of available records and aggregate them into one unique observation by carrier-route for each quarter. In doing this, these records would be aggregated to a few thousand records that were used for analysis. However, Verlinda used one quarter of the government data to demonstrate that the data could be analyzed disaggregately without collapsing the data into average carrier-route observations (12). In doing so, a positive relationship between market concentration and price dispersion is found. Another study using government data also finds a positive relationship between market concentration and price dispersion, although the change in relation is attributed not to the aggregated method of analysis, but to omitted-variable bias in other studies, which the authors correct using an instrumental variables approach (13).

Yet another conflicting finding is that the relationship between market concentration and price dispersion is not strictly positive or negative, but is non-monotonic, inverse U-shaped (14). The authors of this study provide a theoretical model, as well as an empirical model using government data sources, to demonstrate the non-monotonic relationship. In this model, an increase in market concentration when the market is already competitive will result in higher price dispersion while an increase in market concentration when the market is already concentrated enough will result in lower price dispersion.

As seen from this literature, there are many conflicting theories related to airline price dispersion, and the method of analysis greatly influences the findings. One reason why there are so many conflicting theories of price dispersion is the data that is being used. Government data sources for airfares are considered aggregate data in that they summarize and/or randomly sample a small portion of all tickets sold. However, with the widespread use of the internet for booking tickets, there is an opportunity to collect much more detailed and disaggregate data. The use of disaggregate data can be used to

resolve some of these conflicting theories. To date, there have been three studies of price dispersion using disaggregate data. However, two of these studies are for international markets that are not comparable to U.S. domestic markets (15, 16). The other study was analyzed on 12 routes and found a negative relationship between market concentration and price dispersion (17). It is also important to point out that ticket observations used in the price dispersion literature differ across studies; some studies observe actual ticket purchases, while other studies observe offered tickets that may or may not have actually been purchased.

IV. METHODOLOGY

Given the data limitations of past studies that used disaggregate data to study airfare prices, the goal of this research is to study how certain characteristics of a market affect airfares by using a large sample of U.S. domestic markets that cover a broad range of market structures. To the best of the authors' knowledge, the dataset used in this study represents the largest and most comprehensive disaggregate airline pricing database available to the research community.

A. Data

The data was collected in the fall of 2007 in collaboration with QL2 Software®, one of the major U.S. companies that collects competitive pricing and product information from websites. In order to obtain data for Southwest Airlines, additional webbots were written by an academic team at Georgia Tech in order to supplement the data provided by QL2. The data collected consists of prices for more than 100 U.S. markets for one month of departure dates, which were selected to represent periods of peak and off-peak demands (*i.e.*, Thanksgiving and early December 2007). Round-trip and one-way fares were recorded daily for at least 30 days prior to flight departure. Non-stop fares were obtained from each airline's website, while non-stop and connecting fares were obtained from at least one major on-line travel agency (Orbitz®, Travelocity® or Expedia®). For an especially detailed explanation of the data collection methodology and compilation, as well as a more specific account of the dataset, the reader is referred to (18).

A subset of the aforementioned dataset was chosen for data analysis in order to represent a wide variety of interesting market competition structures and airline competition effects, such as monopolies, duopolies, and competitive markets broken down into subcategories representing whether the markets have multi-airport effects and/or LCC presence. In defining these categories of market structures, only airlines that fly non-stop in a market were considered, thus all observations for connecting flights were eliminated. This is in following with the methodology of a number of other researchers (8, 10, 12, 13, 14, 15, 16, 19) and is done for two reasons. Firstly, this ensures that the analysis is somewhat comparable to those of past studies. Secondly, eliminating connecting tickets makes the analysis far less complicated. This is due to the fact that connecting tickets represent significantly different qualities of service than direct tickets and controlling for the cost differences would be more

complex. Additionally, only the lowest fares are included in the observations, which is also comparable to the methodology of other researchers (20, 21, 22). Using the lowest observed fare controls for vertical price differentiation. Vertical price differentiation is the difference in prices due to the differing qualities of tickets (such as restricted vs. non-restricted tickets). By controlling for vertical differentiation, the analysis focuses on horizontal price differentiation, which is defined as the difference in prices due to the varying tastes of customers (such as brand preference, aircraft preference, etc). By focusing on the horizontal price differentiation, the analysis can capture the competitive impacts associated with price dispersion.

The final dataset that was used in this study consists of 108,632 observations for 62 airport-to-airport markets and 12 airlines. Each observation represents the lowest non-stop, round-trip fare that was offered by each airline flying non-stop in the market on the date that the website was queried and for each specific day of flight departure, assuming a one night stay.

B. Analysis of Data

When analyzing the data, it was apparent that any level of aggregation only served to hide some of the most interesting observations in the dataset. Airline pricing policies in the dataset vary greatly by airline, market, peak/off peak time periods, and the number of days from departure. Thus, it was inappropriate to analyze the data in a way that would aggregate any of these variables. Because of this, a case study approach was taken instead of a regression type approach.

V. CASE STUDIES

Several case studies were investigated in this research, including advance purchase restrictions, business vs. leisure markets, markets with codeshares, monopoly markets, and markets with competition from two low cost carriers. For space considerations, only one case study is presented in detail in this paper. However, the sections on public policy and future research include results from the other case studies where appropriate.

A. The Case of Monopoly Markets

In the sample of 62 markets, three types of monopoly markets existed: 1.) One major carrier flies non-stop in the market with no apparent multi-airport effects, 2.) One major carrier flies non-stop in the market with observable multi-airport effects, 3.) One LCC flies non-stop in the market. Each type of monopoly exhibits very different price dispersions, average prices, and carrier pricing strategies.

What is interesting is that out of all of the different market structures, the two most extreme cases on both sides of the spectrum are both monopoly cases. More specifically, monopolies with one major carrier and no multi-airport effects exhibit the highest fares and highest price dispersion out of the entire sample of 62 markets. For these markets, the pricing is very dynamic and fluctuates daily.

On the other hand, monopolies with a LCC as the only non-stop competitor exhibit the lowest prices and lowest price

dispersion out of the entire sample. In these markets, the price dispersion as the date of flight departure approaches stays very flat or decreases slightly for both peak and off-peak periods. This is truly an anomaly for a monopoly market where higher average prices could be charged. Three of the markets analyzed for this type of monopoly were short-haul markets flown by Southwest and exhibit extremely flat pricing. The flat pricing on these markets could be due to Southwest's business model, or could also be due to the fact that they are all short-haul markets. Two of the markets were long-haul markets flown by Air Tran. In these two markets, the average prices are higher and the price dispersion is more dynamic as the day of departure approaches.

For the case in the middle of these two extremes, monopolies with one major carrier and multi-airport effects, the fares are usually lower than those of major carrier monopolies without multi-airport effects. The price dispersion is also lower for most of the markets, with a couple of exceptions.

Another interesting note is that for monopolies with one major carrier (with or without multi-airport effects) the price dispersion trends are different for the peak and off-peak periods. However, for monopolies with a LCC, the peak and off-peak periods have similar trends.

VI. IMPLICATIONS FOR PUBLIC POLICY

In this study, observations were made using disaggregate data on a sample of 62 markets that cover a broad range of market structures. Some of the most important points that were shown include the following:

- Low price dispersion can be associated with both low and high market concentration, depending on the characteristics of the market. This finding contrasts with the findings of past research on price dispersion.
- When a low cost carrier is the only airline competing non-stop on a route, the monopoly route functions much differently than a monopoly with a major carrier. Even in a monopoly situation, low cost carriers (especially Southwest) demonstrate flat pricing and price dispersion as the day of departure approaches.
- As the day of departure approaches, price dispersion can either increase or decrease.
- Peak and off-peak periods often demonstrate different pricing strategies, highlighting the importance of jointly examining price and demand.
- Major carriers tend to exhibit more dynamic pricing strategies than those of low cost carriers, suggesting the former are targeting both business and leisure customers.
- Markets with codeshares (specifically codeshares between US and UA) sometimes exhibit very unusual and high price dispersion on the airline that is selling tickets for a flight operated by another airline.

The results of this study could be used to support analysis of mergers and acquisitions, allocation of gate slots for new entrants, and other policies that relate to airline competition and the assessment of consumer welfare benefits. For example, this paper has shown that there are certain instances when

monopoly routes exhibit lower price dispersion and lower average prices than competitive routes, as is the case of monopoly routes with one non-stop low cost carrier. These differences in monopoly routes highlight the importance of understanding price dispersion at the detailed, disaggregate level when analyzing the impact of future mergers and acquisitions.

Perhaps most importantly from a public policy perspective, this paper shows the importance of disaggregate data that describe individual airline behavior. Much of public policy discussion and analysis relies on average market values that can hide important market behavior. With the advent of internet-based ticketing, a powerful tool now exists that can be used to understand some of the finer detail of airline markets and competition.

VII. FUTURE RESEARCH

In future research efforts, there is a need for disaggregate demand data in order to link pricing strategies with demand as the day of departure approaches. This could be accomplished by pulling seat maps off of the internet while collecting airfares on-line. There is also a need for more research, at the disaggregate level, on how codesharing affects pricing within a market. As more and more airlines begin to use codeshares, understanding the impacts on the market will become more important. There is also a need to link price dispersion to individual revenue management practices of airlines, as there appears to be evidence of more price dispersion in airlines with complex revenue management systems. Additionally, in future research efforts, it would be helpful to compare the offered ticket observations with an actual ticket sample to see which fares were actually purchased. In doing this, market sizes, carrier shares and average fares for each carrier could also be obtained from the ticket sample.

ACKNOWLEDGMENT

This research was supported by the 2008-2009 Graduate Research Award Program on Public-Sector Aviation Issues. We would like to thank the program mentors, John W. Fischer of Congressional Research Service and Richard Golaszewski of GRA Incorporated, for the beneficial comments and feedback. We would also like to thank the program manager, Lawrence Goldstein for his help. Part of this research was also supported by a Dwight D. Eisenhower Graduate Transportation Fellowship. We would like to thank Lauren Bankston and Paul Campbell of QL2 Software for providing the data, along with Angshuman Guin and John Leonard of Georgia Tech and Shawn Pope of Cambridge Systematics for their efforts in data compilation.

Note: A version of this paper, which includes all of the case studies, is currently being considered for publication in Transportation Research Record and is available on the Transportation Research Board 2010 Annual Meeting CD-ROM.

REFERENCES

- [1] Cento, A. (2009). *The Airline Industry: Challenges in the 21st Century*. Physica-Verlag Heidelberg.
- [2] Steenland, D. M. (May 14, 2008). "Hearing on: Impact of Consolidation on the Aviation Industry, with a focus on the Proposed Merger between Delta Air Lines and Northwest Airlines." Testimony to the House Committee on Transportation and Infrastructure, Subcommittee on Aviation.
- [3] Harteveldt, H. H., Wilson, C. P., et al. (2004). "Why Leisure Travelers Book at Their Favorite Site." *Forrester Research: Trends*.
- [4] PhoCusWright (2004). "The PhoCusWright Consumer Travel Trends Survey." PhoCusWright. Downloaded from <http://store.Phocuswright.com/phocrttsusi.html>.
- [5] Borenstein, S. (2005). "U.S. Domestic Airline Pricing, 1995-2004." SSRN.
- [6] Bureau of Transportation Statistics. "First-Quarter 2009 Airline Financial Data: Network Airlines Report Sixth Consecutive Quarterly Loss Margin." www.bts.gov/press_releases/2009/bts030_09/html/bts030_09.html.
- [7] 7. Borenstein, S. (1989). "Hubs and High Fares: Dominance and Market Power in the U.S. Airline Industry." *The RAND Journal of Economics*, 20(3), 344-365.
- [8] Borenstein, S., and Rose, N. L. (1994). "Competition and Price Dispersion in the U.S. Airline Industry." *The Journal of Political Economy*, 102(4), 653-683.
- [9] Hayes, K. J., and Ross, L. B. (1998). "Is Airline Price Dispersion the Result of Careful Planning or Competitive Forces?" *Review of Industrial Organization*, 13(5), 523-541.
- [10] Verlinda, J. A., and Lane, L. (2004). "The Effect of the Internet on Pricing in the Airline Industry." SSRN
- [11] Dana, J. D., Jr. (1999). "Equilibrium Price Dispersion under Demand Uncertainty: The Roles of Costly Capacity and Market Structure." *The RAND Journal of Economics*, 30(4), 632-660.
- [12] Verlinda, J. A. (2005). "The Effect of Market Structure on the Empirical Distribution of Airline Fares." SSRN.
- [13] Gerardi, K., and Shapiro, A. H. (2007). "The Effects of Competition on Price Dispersion in the Airline Industry: A Panel Analysis." SSRN.
- [14] Liu, Q., and Serfes, K. (2006). "Second-Degree Price Discrimination and Price Dispersion: The Case of the U.S. Airline Industry." SSRN.
- [15] Giaume, S., and Guillou, S. (2004). "Price discrimination and concentration in European airline markets." *Journal of Air Transport Management*, 10(5), 305-310.
- [16] Bilotkach, V. (2005). "Understanding Price Dispersion in the Airline Industry: Capacity Constraints and Consumer Heterogeneity." SSRN.
- [17] Stavins, J. (2001). "Price Discrimination in the Airline Market: The Effect of Market Concentration." *The Review of Economics and Statistics*, 83(1), 200-202.
- [18] Pope, S., Garrow, L.A., et al. (2009). "A Conceptual Framework for Collecting Online Airline Pricing Data: Challenges, Opportunities, and Preliminary Results." In *TRB 88th Annual Meeting Compendium of Papers DVD*, Paper 09-1579. (Publication in TRR forthcoming).
- [19] Bilotkach, V., Talavera, O., Gorodnichenko, Y., and Zubenko, I. (2006). "Are Airlines' Price Setting Strategies Different?" SSRN.
- [20] Mentzer, M. S. (2000). "The Impact of Discount Airlines on Domestic Fares in Canada." *Transportation Journal*, 39(4), 35-42.
- [21] Pels, E., and Rietveld, P. (2004). "Airline pricing behaviour in the London-Paris market." *Journal of Air Transport Management*, 10(4), 277-281.
- [22] Bilotkach, V., and Pejcinovska, M. (2007). "Distribution of Airline Tickets: A Tale of Two Market Structures." SSRN.

Evaluating Light Detection and Ranging (LIDAR) Technology in the Terminal Aerodrome Environment for Potential Enhancements and Air Traffic Management

Armond E. Sinclair

Department of Aviation

The Ohio State University

Columbus, OH, U.S.A

Armond.sinclair@gmail.com

Abstract—It is believed that implementing a LIDAR ground based system into a terminal aerodrome will enhance air traffic management and airport operations. Implementing a technology of this magnitude to become a fully marketable system will require additional research and time to better understand the operations of LIDAR and its capabilities in the terminal aerodrome environment.

Keywords- Lidar; Radar; airport; aerodrome; Nexgen

I. INTRODUCTION

As aviation continues to progress into the twenty first century, technological advancements in air traffic management are steadily being integrated into the system. Unfortunately these technological advancements are slow to be implemented due to regulatory and cost constraints. One of the many advanced technologies that have been over looked in airfield operations is the use of LiDar technology. Comparing the efficiencies, accuracies and practicalities of LiDar technology to the current ground based radar systems may prove that airports are operating below their potential efficiency levels. This comparison may also prove that the current radar systems are outdated and can be out performed by LiDar technology. The goal of this research is to study and evaluate the accuracies of LiDar technology versus radar technology in the local terminal aerodrome to study whether such technologies would be an improvement over the current ground based radar technology.

II. BACKGROUND

Throughout its history Air Traffic Control (ATC) has used ground based radar to track aircraft within terminal aerodromes. Their ground based radar systems transmits radio waves outward, which then are reflected back to the antenna after hitting an object, which in this case would be an aircraft. The time it takes a radio wave to be received after transmission will determine the distance of an aircraft from the point of reference. Unfortunately, the clarity of objects plotted on the ATC radar display are two dimensional and limits the controller's capabilities to only detection and ranging. The

ability to decipher or identify the various categories of aircraft is limited to the size of the radar return signature. If an aircraft were to fly into an airspace without a transponder, ATC would not be able to properly identify the aircraft until it appears within the field of view (FOV). For safety and security reasons, these methods of operation are outdated, inefficient, and unsafe. The implementation of Light Detection and Ranging (LiDar) technology would improve deficiencies that currently exist at airports worldwide.

LiDar operates very similar to radar by transmitting through transmitters and receiving through receivers. It uses much shorter wavelengths of the electromagnetic spectrum, typically in the ultraviolet, visible, or near infrared range. A laser typically has a very narrow beam which allows the mapping of physical features with very high resolution compared to radar. The significance of this technology is accredited to LiDar's ability to produce an exponentially increased accuracy of detection, ranging, and high resolution three dimensional mapping by providing not only X and Y axis points but also Z axis points as well. Traditionally LiDar has been used for topographical, meteorological, and astronomical applications because of its accurate three dimensional modeling and ability to detect objects that radar cannot. For example, LiDar has the ability to detect small objects such as birds, which can significantly reduce bird strikes, resulting in the safety of lives and multimillion dollar assets.

Many companies use LiDar technology in the form of airborne laser scanning, which is a top down approach of scanning terrain to acquire an accurate three dimensional portrait of the desired space. Taking these same fundamental concepts and applying it to airfields would increase safety, security, and operations. Instead of using a top down approach, LiDar technology can be used in a bottom up approach. Implementing LiDar in this capacity would allow the user to scan the airspace and accurately view the entire airfield in a three dimensional perspective. This capability would enable the controller to view the airfield from infinite angles, increasing situational awareness, safety of aircraft in the vicinity, security of the airspace, and flow of operations. The continuous

scanning of LiDar provides up to date real time data in a smooth uninterrupted flow.

Fortunately the technology for the hardware and software currently exist and has already been implemented on autonomous vehicles. Test results prove that not only does LiDar provide a real time high resolution three dimensional display, but it also provides the user the ability to adjust the view to suit his or her needs. The HDL-64E, for example, is a LiDar system that operates on a rather simple premise:

Instead of a single laser firing through a rotating mirror, 64 lasers are mounted on upper and lower blocks of 32 lasers each and the entire unit spins. This design allows for 64 separate lasers to each fire thousands of times per second, providing far more data points per second and a much richer point cloud than conventional designs. The 64 lasers are employed with each laser/detector pair precisely aligned at predetermined vertical angles, resulting in a 26.8 degree vertical FOV. By spinning the entire unit at speeds up to 900RPM (15 Hz), a 360 degree FOV is inherently delivered. Regardless of the spin rate, over 1.3 million data points (i.e. pixels) are generated each second, providing an exponentially richer point cloud than ever before possible. The HDL-64E supplies returns out to 120 meters. It features ~1 inch distance accuracy and excellent repeatability. Radial resolution is dictated by spin rate, with radial accuracy as precise as .05 degrees. Additionally, state-of-the-art signal processing and waveform analysis are employed to provide high accuracy, extended distance sensing and intensity data. –Velodyne [1]

The HDL-64E S2 outputs UDP Ethernet packets. Each packet contains a data payload of 1206 bytes that consists of 12 blocks of 100-byte firing data followed by six bytes at the end of each packet that contains a spin counter and firmware version information. Each packet can be for either the upper or lower laser banks (called “laser blocks”) - each bank contains 32 lasers.

The packet format is as follows:

- 2 bytes of header info: This header indicates whether the packet is for the upper block or the lower block. The upper block will have a header of 0xEEFF and the lower block will have a header of 0xDDFF.
- 2 bytes of rotational info: This is an integer between 0 and 35999. Divide this number by 100 to get degrees from 0. 32 laser returns broken into 3 bytes each. Each return contains two bytes of distance information in .2-centimeter [2mm] increments, and one byte of intensity information (0 – 255, with 255 being the most intense return). A zero distance value within the data packet indicates there are no returns up to 120 meters, the maximum range of the device. Six status bytes at the end of the packet: 2 bytes spin count (binary). This field is incremented for each revolution. After 65,535 revolutions, the counter resets to 0. 4 bytes alternating between:

- A reading showing the internal temperature of the unit: You will see a " DegC" ASCII string as the last four bytes of the packet. The two bytes before this string are the thermistor's reading in C in hex 8.8 format. This is in "big endian format" - i.e. the byte immediately preceding the DegC text is the whole degrees, and the byte preceding that is the fraction of a degree in 1/256 increments. So if you see c0 1a, the temperature of the thermistor is 26.75 degrees C. The version number of the firmware in ASCII character format " Vxxx" where “xxx” is the version number, e.g. "25b" which represents version 2.5b (the most current software version as of this writing).

The minimum return distance for the HDL-64E S2 is approximately 3 feet [.9 meters]. Returns closer than this should be ignored. The HDL-64E S2 data is presented as distances and intensities only. Velodyne includes a packet viewer called DSR, whose installer files are on the CD that came with the unit. DSR reads in the packets from the HDL-64E S2 unit, performs the necessary calculations to plot the points presented in 3-D space, and plots the points on the viewer screen.

Each HDL-64E S2 laser is fixed with respect to vertical angle and offset to the rotational index data provided in each packet. For each data point issued by the HDL-64E S2, rotational and horizontal correction factors must be applied to determine the point's location in 3-D space referred to by the return. Each HDL-64E S2 unit comes with its own unique .XML file, called db.XML, that was generated as a result of the calibration performed at Velodyne's factory. DSR uses this XML file to display points accurately. The .XML file also holds the key to interpreting the packet data for users that wish to create their own software applications. db.XML contains 64 instances of the following five values used to interpret the packet data:

- rotCorrection: This parameter is the rotational correction angle, in degrees, for each laser, as viewed from the back of the unit. Positive factors rotate to the left, and negative values rotate to the right.
- vertCorrection: This parameter is the vertical correction angle, in degrees, for each laser, as viewed from the back of the unit. Positive values have the laser pointing up, and negative values have the laser pointing down.
- distCorrection: Each laser has its own unique distance due to minor variations in the parts used to construct the laser. This correction factor, in centimeters, accounts for this variance. This number should be directly added to the distance value read in the packet.
- vertOffsetCorrection: This value represents the height of each laser, in centimeters, as measured from the bottom of the base. It is a fixed value for all upper block lasers and a different fixed value for all lower block lasers.
- horizOffsetCorrection: This value represents the horizontal offset of each laser, in centimeters, as

viewed from the back of the laser. It is a constant positive or negative value for all lasers.

Taking the same concepts of the HDL-64E S2 and scaling it up to suit the needs of a terminal aerodrome will create an enhanced environment that promotes efficiency and safety. One of the advantages of this technology is that multiple units can be integrated together to make one large image. To maximize optimal viewing for ATC, four units would be installed at an airfield limiting possible blind spots. Each unit has 360 degree FOV, but once an object reflects the light, a shadow is casted behind the object. To help reduce casted shadows, the multiple units will overlap and only depict the object that is reflecting the light.

When speaking to Mr. Deral Carson, Air Traffic Manager for Midwest Air Traffic Services Inc., at The Ohio State University Don Scott airfield, he spoke about the multiple advantages that ground based LiDar systems would have over ground based radar systems within the terminal aerodrome. The first advantage he noticed after seeing the software was the ability to identify what you were looking at visually. He further stated that visual recognition would help ATC understand the performance and limitations of aircraft without even speaking with the crew. This capability would reduce radio chatter and promote a more efficient dialog with the aircrews. The second advantage that was noticed by Mr. Carson was the rate at which the LiDar system updates its information. Typically, ground based radar systems only update once every six seconds. The LiDar system, which we are currently using, updates continuously with a real time streaming display. There are no breaks between cloud points. The third advantage that was noticed was LiDar's ability to distinguish altitudes. Having the capability to determine an aircraft's altitude without a transponder is a phenomenal technological advancement, and would significantly aide ATC in IFE scenarios. When asked to provide any final statements about the LIDAR system, Mr. Carson said, "The speed at which LIDAR operates is mind blowing". He further stated that "Typically the FAA is always looking for replacement systems and since nextgen doesn't solve all the issues something like this could be really advantageous".

One of the concerns that currently exist in regards to LiDar technology is the regulations on the use of lasers where there is a high concentration of air traffic. The FAA has the authority to regulate the safe and efficient use of the navigable airspace (Title 49 U.S.C., Section 40103, Sovereignty and Use of Airspace, and the Public Right of Transit) [2], and the Food and Drug Administration (FDA) has the authority to regulate light emitting products and electronic product radiation (Title 21, U.S.C. Section 360hh.) [3]. To meet the existing requirements currently in place by government agencies for the use of LiDar technology in airfield operations, the type of laser waveform used must be a Class 1 non-visible eye safe laser. Research has shown that this type of waveform has detection and ranging capabilities up to 20-30 kilometers, which is sufficient enough to cover any Class B Airspace.

Some may speculate whether the investment of LiDar technology in airfield operations is worthy of serious consideration because of NextGen and ADS-B. LiDar is not a suggestion to replace NextGen but rather a technology that can

compliment it. Unfortunately, one of the drawbacks of NextGen and ADS-B is that it relies on GPS technology and is susceptible to the effects of space weather. According to a research study performed at Cornell University, "strong solar flares cause Global Positioning System (GPS) receivers to fail. Such failures could be devastating for 'safety-of-life' GPS operations -- such as navigating passenger jets." [4] The U.S. Department of Transportation has not only recognized space weather as a factor that causes interruptions in GPS operations, but have also stated that intentional loss of GPS connections can come from "intentional interference" which include jamming and spoofing counterfeit signals. [5] Airports such as JFK cannot afford to have ADS-B failures on final approach. A 15-30sec loss of GPS communications due to phase misalignment or carrier signal cycle slips can present a platform for destruction. Having a ground based LiDar system would enhance NextGen what would be a more accurate and surveillance technology in the terminal area.

This research will analyze the potential benefits of applying LiDar technology to the terminal aerodrome air traffic environment through a multifaceted program of empirical analysis of applying data from the analysis into statistical models and performing benefit cost analysis of LiDar implementation versus the current ground based radar infrastructure.

The empirical analysis will consist of field testing the use of LiDar equipment at The Ohio State University Airport, will include clarity of resolution, position accuracy on the X,Y,Z plane, sampling rates, and other factors which contribute to the sequencing of aircraft on approach to an airfield. The data collected will be applied to model to compare to determine if the capabilities contribute to a more enhanced air traffic flow. An evaluation will be made of the cost of the implementation in comparison to any potential benefits.

III. SUMMARY

Implementing ground based LiDar technology is hypothesized to enhance the flow and efficiencies of terminal aerodrome operations by providing enhanced aircraft detection and ranging capabilities, accurate aircraft identification capabilities, three dimensional ATC monitoring, and an increased detection of airfield hazards. This short paper describes the early stages of background research. Empirical research and modeling will be conducted over the next twelve months.

REFERENCES

- [1] Velodyne LIDAR Inc, Morgan Hill, CA 95037 USA
- [2] Code of Federal Regulations, Title 49 U.S.C., Section 40103, *Sovereignty and Use of Airspace, and the Public Right of Transit*, Washington D.C. USA
- [3] Code of Federal Regulations, Title 21, U.S.C. Section 360hh, *Radiation Control Program*, Washington D.C. USA
- [4] Carroll, James "Vulnerability Assessment of the Transportation Infrastructure Relying on GPS", U.S. Department of Transportation Volpe Center Power Point Presentation, October, 5th 2001
- [5] Dubey, S., Wahi R., and A.K Gwal, "Effect of Ionospheric Scintillation on GPS Receiver at Equatorial Anomaly Region Bhopal" Working Paper, Barkatullah University, Bhopal, India

- [6] Federal Aviation Administration. *Advisory Circular 150/5200-37 - Introduction to Safety Management Systems for Airport Operators*. Retrieved February 12, 2009, from http://www.faa.gov/airports_airtraffic/airports/resources/advisory_circulars/media/150-5200-37/150_5200_37.pdf
- [7] Federal Aviation Regulations, Part 139 "Certification of Airports". US Code of Federal Regulations, Washington, DC.
- [8] ACRP Report 1, *Safety Management Systems for Airports, Volume 1: Overview*, Transportation Research Board, Washington, D.C. 2007
- [9] Federal Aviation Administration. *Federal Aviation Administration Safety Management System Manual*. Retrieved March 24, 2008, from http://platinum.ts.odu.edu/Apps/FAAADCA.nsf/FAA_SMS_Manual-Version_1_1.pdf

The Aircraft Sequencing Problem

Nadiah Ahmad

Department of Integrated Systems Engineering
The Ohio State University
Columbus, USA
ahmad.87@osu.edu

Abstract— This paper describes a research on the aircraft sequencing problem intended for a Master thesis. The aircraft sequencing problem is a problem of determining the landing times for a sequence of aircrafts on available runways in a way that they land during the predetermined time windows for landing while respecting separation criteria for safety. The goal of this research is to model and analyze the problem from several viewpoints to determine the criteria for optimization. This will involve formulation of appropriate objective functions and comparison of computational results after solving the problem using methods such as mixed integer programming and network flow programming.

Keywords-aircraft sequencing; scheduling; air traffic control

I. INTRODUCTION

The aviation industry has experienced a substantial growth in air traffic volume over the last decade and air traffic volume is expected to continuously increase in the future. In the US alone, despite declines during the recession, air traffic is forecasted to grow on average from 3 to 4 percent per year. By 2025, the Federal Aviation Administration (FAA) has forecasted about 1.7 trillion available seat miles (ASMs) to be flown by the US commercial airlines with a total of 1.1 billion passenger enplanements [1]. This significant growth in the air traffic is a major concern to the air traffic management (ATM) as they have to deal with traffic congestions that arise from it.

Physical expansions and new airport constructions to accommodate the growing air traffic demand are very limited due to geographical and environmental constraints [2]. Consequently, some airports have become very congested, especially the hub airports. At these airports, the air traffic demands during certain period of times are much higher than the capacities that the airports can handle. This demand capacity imbalance can create delays not only to the operations at the affected airports but also to the overall national airspace system (NAS), especially when the affected airports are the hub for connecting airlines [3].

One important step to deal with the congestion at an airport is to optimize the current operation at the airport terminal by making use of its capacity efficiently. For many airports, the runways are seen as the bottleneck of the system as their capacities depend on the runway throughputs. All aircrafts arriving and departing at or from the same airport have to be merged at the available runways. At the same time, these aircrafts must meet the minimum separation limit imposed

between successive aircrafts in order to maintain the safety level in the air traffic.

Inefficient aircraft sequencing can cause many problems such as extended delays, missing minimum separation requirement, inefficient fleet mix and timeslot use by small aircraft [3]. It is therefore crucial for the air traffic controller who is responsible for directing and coordinating aircrafts in the space and on the ground to optimize the number of aircrafts that could land or depart from runways through efficient aircraft sequencing or scheduling. In the sequencing, aircrafts must respect the separation requirement based on the wake vortex criteria.

In most airports however, the air traffic controllers assign landing time to a set of aircrafts based on the first-come first-served (FCFS) principle which may result in inefficient sequencing and scheduling. Furthermore, this problem can also increase operational cost for airlines and disrupt the safe separation between aircrafts [3]. These situations serve as a motivation for our research which aims to explore the modeling of the aircraft sequencing problem (ASP) from diverse different perspectives and conduct some computational analysis intended for a Master thesis. The focus of this work will be for a single runway problem and may be extended to multiple runways problem. Currently, our work is still at the beginning and therefore we do not provide much progress in this paper. However, in the next sections, we will describe the aircraft sequencing problem in section II, some literature reviews in section III, the ASP formulation in section IV and our research goals in section V.

II. AIRCRAFT SEQUENCING PROBLEM

In the ATC segment, a traffic management adviser (TMA) assigns fix-crossing times to a set of aircrafts trying to land at the same airport. The assigned times are called the scheduled time of arrival (STA). These aircrafts are sequenced in such a way that they arrive at the approach fix during the STA by using speed control, vectors, etc.

The ASP thus involves determining the landing times for a sequence of aircrafts on available runways in a way that they land during the predetermined time windows for landing while respecting separation criteria for safety. An earliest possible arrival time at the fix and the latest possible arrival time bound the predetermined time window for landing [4]. The earliest time is normally the time the aircraft approach the fix if it flies at a maximum speed or through the shortest route. The latest

time, on the other hand is the approach time until the fuel runs out if the aircraft is put into a holding pattern [2].

In a single runway problem, a set of n aircrafts have different preferred landing times. The controller must find a sequence of aircrafts that optimizes a certain measure of performance by satisfying all the constraints. Aircraft preferred landing time is determined based on most fuel-efficient speed called the cruise speed. Flying at other than cruise speed will incur costs to airlines (i.e. due to holding factors or separation assurance). These costs normally grow as the deviation from the preferred landing time grows.

The separation between successive aircrafts must be greater than the minimum specified. This separation is specified by the landing separation time which depends on the type of aircrafts. The separation time required when a small aircraft follows a big aircraft is bigger compared to when a bigger aircraft tails a small aircraft. Therefore, different sequences may produce different runway throughput.

In the ASP, one important aspect when determining the aircraft sequence is measuring the quality of the sequence. Fahle et al. [2] point out there are at least three parties that have different objectives with regards to the aircraft scheduling. First, the ATC who is responsible for deciding the actual sequence have the objectives of maximizing the runway throughput, minimizing airborne delays, and minimizing the controllers' workload. The airlines on the other hand, are concerned with the punctuality to schedule, maximizing incoming and outgoing flight connectivity as well as respecting the priority of one flight over the others. Lastly, the airport has the goals of maximizing operation punctuality and avoiding the need for gate changes caused by delays. Given the various aims involved, Fahle et al.[2] however highlight that punctuality is the main aspect of a good schedule even though it can be defined differently by different parties.

III. LITERATURE OVERVIEW

The ASP is known as an NP hard problem [5]. This means that there is no efficient algorithm that can solve the problem optimally in polynomial time as the problem gets larger. In the past, this problem has been tackled using exact methods, heuristics, dynamic programming and simulation. Some early work on ASP can be seen in [4][6][7]. These works are based on mixed integer problem (MIP) formulations. Beasley et al. [4] solve the problem using LP-based tree search and branch and bound algorithm. Extensive literature review on ASP can also be referred from [4]. Hu and Poalo [5] use the application of genetic algorithms (GAs) to tackle the ASP in multi-runway systems. They show that the solution produced exhibits a good potential of real-time implementation in the ASP. Xianbin et al [8] use Cellular automation and GA that can generate feasible solution fast and better than FCFS algorithm.

Xuezhi and Shuang [9] model the problem as a one machine scheduling problem with setup dependent time and introduce a multifactor decision making model with every aircraft having weighted attribute. By optimizing the weighted total completion time of landing, the problem is solved with greedy algorithm and tested on 10 planes. As most of published work suggest more complex sequencing algorithms compared

to traditional FCFS method, Brentnall and Cheng [10] argue that there is a need to study the effect of using those suggested algorithms on ATC system. Therefore, a comparison study was done using simulation models based on four selected sequencing algorithms, which concludes that FCFS is a robust method in some situation.

IV. ASP FORMULATION

This section presents the MIP formulation of the ASP problem described in [4] and [2]. The formulation presented here only considers the problem for the single runway. The notation involved in this formulation is as follows:

$$P = \text{the set of aircrafts}$$

$$E_i = \text{the earliest landing time of aircraft } i \in P$$

$$L_i = \text{the latest landing time of aircraft } i \in P$$

$$S_{ij} = \text{the minimum separation time required between aircraft } i \text{ and } j \text{ if aircraft } i \text{ lands before aircraft } j \text{ for } \forall i, j \in P$$

$$T_i = \text{the target or preferred time for landing of aircraft } i \in P$$

$$A_i = \text{the penalty cost per unit time for aircraft } i \in P \text{ when landing before its target time } T_i$$

$$B_i = \text{the penalty cost per unit of time for aircraft } i \in P \text{ when landing after its target time } T_i$$

The variables are:

$$t_i = \text{the landing time of plane } i \in P$$

$$d_{ij} = 1 \text{ if aircraft } i \text{ land before aircraft } j \text{ and 0 otherwise } \forall i, j \in P, i \neq j$$

$$a_i = \text{the duration aircraft } i \in P \text{ lands before its target time } T_i$$

$$b_i = \text{the duration aircraft } i \in P \text{ lands after its target time } T_i$$

Constraints are as follows:

$$E_i \leq t_i \leq L_i \quad \forall i \in P \quad (1)$$

$$d_{ij} \in \{0,1\} \quad \forall i, j \in P, i \neq j \quad (2)$$

$$d_{ij} + d_{ji} = 1 \quad \forall i, j \in P, i < j \quad (3)$$

$$t_i + S_{ij} \leq t_j + M d_{ij} \quad \forall i, j \in P, i \neq j \quad (4)$$

$$t_i + b_i - a_i = T_i \quad (5)$$

Constraint (1) ensures that all aircrafts land within their respective time windows. The order in which the aircrafts land is satisfied by conditions (2) and (3). Constraint (4) is for satisfying the separation constraints where M is a large number. Lastly, constraint (5) relates the deviation from the target time to the actual landing time. This condition help create the objective function which is as follows:

$$\text{Min } \sum_{i \in P} (B_i b_i + A_i a_i) \quad (6)$$

V. RESEARCH GOALS

The goal of this research is to model and solve the ASP from several different viewpoints such as from the airport operator's perspective, airlines' perspective or a combination of these. Therefore, our focus will be in modeling and testing various versions of the problem. Several approaches to model and solve this problem will be considered in this research. The scope of this research will focus on the single runway problem and the multiple runways problem. The following will discuss the details:

A. Objectives

In this research, we are interested in modeling the ASP with precedence constraints. In this case, approaching aircrafts travelling on the same air-route are not allowed to overtake the other. The precedence constraints may also come from airline's preferences such that one aircraft must land before the other.

One of our aims is also to investigate how well the problem can be modeled with various objective functions. Most published papers on the ASP adopt minimizing the total cost incurred for deviating from the aircrafts target landing times as the objective function. In some others, the objective is simply minimizing the total completion time of landing or the time of landing for the last aircraft. One consideration for this research is to model the problem when the aircraft priorities are also considered. This means that, some aircrafts may have certain priorities over the others due to factors such as early arrivals or high passenger enplanement. This may require us to explore such factors that affect the flight priorities. Another element that can be considered is when the objective includes minimizing the airborne delays for aircrafts. This can be measured by the period the aircraft enter the holding pattern before it can land.

B. Method

We plan to model and solve our ASP using several approaches and compare their computational results. One approach is to solve the problem using mixed integer programming (MIP) which is the most popular method in the ASP. In addition, network optimization methods will also be considered when appropriate. Previous work has shown that the ASP problem can be modeled as a modified shortest path network problem [11]. The goal is to find the solution to the

problem within reasonable time with good quality solution. It is therefore our interest to investigate the best approach to solve the problem. The computational results will be analyzed and compared across the methods used.

ACKNOWLEDGMENT

The author would like to acknowledge Simge Küçükyavuz and Seth Young for their support and input as advisors for this Master thesis work. The author also wishes to thank them for their reviews and comments during the completion of this manuscript.

REFERENCES

- [1] FAA Aerospace Forecasts FY 2009-2025. January 15, 2010.[Online]. Available:http://www.faa.gov/data_research/aviation/aerospace_forecasts/2009-25/media
- [2] T. Fahle, R. Feldmann, S. Gotz, S. Grothklags, and B. Monien, "The aircraft sequencing problem," *Computer Science in Perspective*, 152-166, 2003.
- [3] L. Le, G. Donohue, and C. Chen, "Auction-based slot allocation for traffic demand management at Hartsfield Atlanta International Airport: A case study", *Transportation Research Record* 1888, no. 1 (1): 50-58, 2004.
- [4] J. E Beasley, M. Krishnamoorthy, Y. M. Sharaiha, and D. Abramson, "Scheduling aircraft landings--the static case," *Transportation Science* 34, no. 2 , 180-197, 2000.
- [5] X. Hu and E. Dipaolo, "An efficient genetic algorithm with uniform crossover for air traffic control," *Computers & Operations Research* 36, no. 1, 245-259, 2009.
- [6] L. Bianco, B. Nicoletti, and S. Ricciardelli, "An algorithm for optimal sequencing of aircraft in the near terminal area," *Optimization Techniques*, 443-453, 1978.
- [7] A. T. Ernst, M. Krishnamoorthy, and R. H. Storer "Heuristic and exact algorithms for scheduling aircraft landings," *Networks* 34, no. 3, 229 – 241, 1999.
- [8] C. Xianbin et al, "A real-time schedule method for aircraft landing scheduling problem based on cellular automaton," in *Proceedings of the 1st ACM/SIGEVO Summit on Genetic and Evolutionary Computation, GEC'09*, Shanghai, China, 2009.
- [9] X. Xuezhi, and C. Shuang, "Researches on Optimal Scheduling Model for Aircraft Landing Problem" In *ICIE '09. WASE International Conference*, 418-421. Vol. 1, 2009.
- [10] A. R. Brentnall and R. C. H. Cheng, "Some effects of aircraft arrival sequence algorithms. *Journal of the Operational Research Society* 60, no. 7 (6),962-972, 2008.
- [11] H. Balakrishnan and B. Chandran, "Scheduling aircraft landings under constrained position shifting," in *AIAA Guidance, Navigation and Control Conference and Exhibit*, Keystone, CO, 2006.

Models for Aircraft Landing Optimization

Mohammad Mesgarpour

School of Mathematics

University of Southampton

Southampton, SO17 1BJ, UK

Email: m.mesgarpour@soton.ac.uk

Chris N. Potts

School of Mathematics

University of Southampton

Southampton, SO17 1BJ, UK

Email: c.n.potts@soton.ac.uk

Julia A. Bennell

School of Management

University of Southampton

Southampton, SO17 1BJ, UK

Email: j.a.bennell@soton.ac.uk

Abstract—Due to an anticipated increase in air traffic during the next decade, air traffic control in busy airports is one of the main challenges confronting controllers in the near future. Since the runway is often a bottleneck in an airport system, there is great interest in optimizing usage of the runway. Our study first presents a brief review of the aircraft landing problem. A model for the problem is then introduced, and possible solution approaches are discussed.

I. INTRODUCTION

Airport capacity (and hence runways and gates) is increasingly becoming a limiting factor in meeting the rising demand for more flights. One of the main factors that determines the runway throughput at airports is the required separation between aircraft during take-off and landing. Dependency of separation on the leading and trailing aircraft and the type of aircraft add to the challenge of solving the sequencing and scheduling problem. Due to its complexity, it is hard to find the optimal solution to the problem in most cases. Thus, it draws significant attention from different scientific communities with numerous research studies carried out on modelling and developing algorithms to increase capacity at an airport.

Since the appearance of a paper of Blumstein [1] on estimating the capacity of an arrivals runway, there have been a variety of studies on airport runway optimization. Although the literature during the last three decades contains more than sixty publications on aircraft landing optimization, the majority of the proposed methods have never been implemented.

The Aircraft Landing Problem (ALP), which is the focus of our work, is to sequence landing aircraft onto the available runways at an airport and to assign each aircraft a landing time, subject to variety of operational constraints. The simple way of sequencing and scheduling of landing aircraft on a single runway is using the First-Come-First-Served (FCFS) approach which assigns scheduled landing time to each aircraft based upon the sequence implied by the earliest time that the aircraft can land. It has been found that FCFS is rarely the best sequencing order in terms of capacity, average delay or even average passenger delay [2].

The prime responsibility of the air traffic controllers is safety of the flights. Standard vertical and horizontal separations which keep flights apart provide one of the main Air Traffic Control (ATC) safety devices. Vortices generated by the aircraft as a consequence of their lift are one of

the reasons for imposing separations. The minimum required separation governs the minimum permissible distance interval between aircraft lined up in sequence on the approach to landing on the runway. Generally, the Wake Vortex (WV) separation required between consecutive aircraft depends on the type of the aircraft, and therefore it is sequence dependent. During peak capacity operations, the WV is often a major concern. It effectively determines runway throughput and thus limits airport capacity in the terminal airspace. Significant asymmetries in minimum required separations can offer an opportunity to reduce airborne delays by shifting aircraft positions in the landing sequence, although finding the best way to achieve this presents a challenge.

There are well-known procedures for making an aircraft wait to land, such as using Vector-for-Space (VFS) or Holding Pattern (HP) [3]. Nevertheless, reassigning an aircraft to landing time far from its initial place in FCFS sequence is not always feasible because of the operational constraints in practice. The *Constrained Position Shifting* (CPS) concept is introduced by Dear [4] for the ALP to prevent the final positions of aircraft in landing sequence from differing from the FCFS order by more than a pre-specified number, called the *Maximum Position Shift* (MPS). Furthermore, when the MPS is small, it maintains fairness among the aircraft by not deviating too far from the FCFS sequence. *Relative Position Shifting* (RPS) is a variant of MPS that takes into account the closeness of the aircraft to the runway when specifying the MPS.

This paper focuses on the techniques and tools of operational research and management science (OR/MS) for solving the ALP. Section II is a brief review of the literature on the use of OR/MS techniques for tackling the ALP. A mixed integer programming (MIP) model is proposed in Section III. Finally, future research directions are discussed in Section IV.

II. LITERATURE REVIEW

Most previous research on the ALP considers a static (off-line) environment based on a given set of aircraft operating over a predefined time horizon. However, a more realistic model considers a dynamic (on-line or real-time) environment, where solutions are revised as aircraft arrive over time and new information becomes available.

A variety of OR/MS techniques have been used to model the problem, such as mixed integer programming (MIP)

and queueing theory, while commonly used approaches for machine scheduling and travelling salesman problems (TSP) are also useful in the development of solution algorithms. Various search techniques have been also applied to solve the problem such as Dynamic Programming (DP), Branch-and-Bound, Branch-and-Price, Genetic Algorithms/Programming, Ant Colony Optimization, etc.

Based on a DP approach for solving the TSP, Psaraftis [5] develops three algorithm for the static case of the ALP to examine two alternative objectives, the landing time of the last aircraft, and the total passenger delay with respect to FCFS discipline. The CPS concept is also incorporated within the algorithm. Bianco et al. [6] point out the relationship between the ALP, a machine scheduling problem (denoted by $1|r_j, \text{seq-dep}|\sum C_j$) with n jobs and the cumulative TSP with ready times. A dynamic programming formulation and three lower bounds are proposed for the scheduling problem.

The scheduling of aircraft landing is formulated as a single machine scheduling problem in Bayen et al. [7]. A DP approach and a linear programming relaxation and rounding are used in the main algorithm. Two approximation algorithms for the sum of arrival times of all aircraft and the arrival time of the last aircraft have performance bounds of 5 and 3, respectively. Because different classes of aircraft are not considered, the required separation between landings is independent of the aircraft type.

Recently, Balakrishnan and Chandran [8] present a DP-based approach to maximize the runway throughput (equivalently, minimizing the landing time of the last aircraft). The problem of scheduling landing aircraft in a CPS environment is considered subject to various operational constraints imposed by arrival time windows, minimum aircraft separation requirements, and precedence relations. The problem is formulated as a modified shortest path problem on a network with $O(n(2k+1)^{2k+2})$ arcs, where n is the number of aircraft and k is the maximum position shift.

Chandran and Balakrishnan [9] also propose a DP algorithm to compute the tradeoff curve between the robustness (reliability of a schedule) and throughput based on their earlier work [8]. More recently, Lee and Balakrishnan [10] extend the previous framework [8], [9] and present a DP algorithm for minimizing the total delay costs of an arrival schedule. Also, the problem of minimizing the fuel costs of the arrival schedule as the main operating cost for most airline, has been studied using the proposed algorithm by allowing the earliest landing time to be less than the estimated landing times which is known as Time Advance (TA). Several polynomial-time DP algorithms for the ALP based on machine scheduling concepts are presented in [11], [12]. Six sequencing algorithms which include three DPs, two FCFS rules, and a heuristic that represents a potential algorithm for an operational AMAN (Arrival Manager) system are implemented. Moreover, four delay sharing strategies include all delay in hold, delay as late as possible, delay as early as possible, and delay evenly throughout the route strategies are implemented.

The literature also contains several branch-and-bound algo-

rithms for the ALP. For example, in addition to an extensive literature overview on the ALP, Beasley et al. [13] develop algorithms for both single and multiple runways. The model is based on an earlier MIP formulation presented in [14].

Different metaheuristic approaches have been examined for scheduling landing aircraft. One of the first and the simplest application of a GA in minimizing the earliness/lateness for the ALP is investigated by Stevens [15]. Based on his work on the permutation-based approach, Ciesielski and Scerri [16], [17] compare two GAs for a real-time dynamic ALP in terms of the percentage of valid solutions and best fitness by specifying a 30-second time slot and variable times between landings. Cheng et al. [18] design four different GAs for solving the multiple runway ALP.

A Population Heuristic (PH) is developed by Beasley et al. [19] to improve the utilization of a single runway. The algorithm aims to minimize the squared deviations from estimated landing time in the presence of five separation criteria. Later, Hansen [20] examines the efficiency and effectiveness of various genetic approaches for the ALP. Regarding the objective function, three different formulations are presented by Capri and Ignaccolo [2] with respect to minimizing the delays which depend on the aircraft classes, maximizing the system capacity, and minimizing the sum of landing times.

The Receding Horizon Control (RHC)-based GA that is introduced by Hu and Chen [21] minimizes the airborne delay, which is the deviation of actual landing time from estimated landing time in a dynamic environment. Hu and Paolo [22], [23] experiment with alternative solution representations in a single and multiple runway ALP. Two different population heuristics, Scatter Search and the Bonomic Algorithm, are applied by Pinol and Beasley [24] to the multiple runway ALP. Both linear and non-linear objective functions are considered.

Ant Colony Optimization (ACO), which is a constructive metaheuristic technique with biological foundation, has been applied to the ALP by Randall [25]. His algorithm aims to minimize the difference between an estimated landing time and the actual landing time for each aircraft, subject to a specified time window and separation criteria.

Dynamic programming exhibits the best performance among exact methods because of its enumerative nature. However, the ability to control run time in local search methods such as GAs makes them serious candidates for use in decision support tools for air traffic controllers.

III. PROBLEM DEFINITION

The majority of research on the ALP considers sequencing the aircraft in the Terminal Manoeuvering Area (TMA). However, sequencing aircraft further away from the airport (such as Extended TMA) may produce better results. The problem is divided into three time stages.

- Stage 1 (*Sequencing Stage*): The first stage starts by entering the aircraft into the airport landing planner's radar range about 40 minutes before touchdown.
- Stage 2 (*Modifying Schedule Stage*): The second stage starts 11 minutes before landing and takes 8 minutes and

includes the final approach step.

- Stage 3 (*Freezing Stage*): The last stage consists of the runway controller's range of operation which is 3 minutes long.

Sequencing and scheduling of arrival aircraft are performed in *stage 1*. As time progresses and new aircraft enter the sequencing stage, the sequence and schedule have to be updated, which is done every five minutes. In *stage 2*, the sequence is not usually changed, with only minor adjustments to the schedules being made. As the aircraft is so close to the runway in *stage 3*, neither the sequence nor the schedule can be modified.

A. Notation

Decision Variables

SLT_i The scheduled landing time of each aircraft i .

X_{ij} Defined to be 1 if aircraft i lands before (not necessarily immediately) aircraft j , and 0 otherwise.

Parameters

n The number of aircraft to be scheduled.

A The set of available aircraft for landing, $A = \{1, \dots, n\}$, which is updated every 5 minutes.

The parameters below are defined for each aircraft i , for $i = 1, \dots, n$.

TLT_i The target (or expected) landing time of aircraft i based on the assigned time slot which is normally specified in flight plan.

Est_i The estimated (or predicted) landing time of aircraft i calculated by trajectory synchronizer equipment after entering the aircraft into the radar range, and is normally based on the FCFS sequence.

Al_i^- The allowed earliness for aircraft i to land.

Al_i^+ The allowed lateness for aircraft i to land.

ELT_i The earliest possible landing time for aircraft i , subject to technical and operational restrictions.

LLT_i The latest possible landing time for aircraft i which is usually determined from fuel limitation, maximum allowed delay, or meeting a connecting flight.

E_i The earliness penalty cost per unit for aircraft i to be advanced more than Al_i^- .

L_i The lateness penalty cost per minute for aircraft i to be delayed more than Al_i^+ .

P_{ij} Defined to be 1 if aircraft i must land before (not necessarily immediately) aircraft j , and 0 otherwise.

S_{ij} The minimum time separation between aircraft i and j , if aircraft i lands before aircraft j .

TS_i The time shifting of aircraft i , which is the maximum time deviation of this aircraft from/to Est_i in the landing sequence.

FB_i^+ Average required fuel burn cost per minute for aircraft i to be delayed.

FB_i^- Average required fuel burn cost per minute for aircraft i to be advanced.

B. Objective Function

Choosing an appropriate objective function for the ALP is controversial and not all stakeholders (ATC, airport, airlines, and government) agree on the selection process. However, the following multi-criteria objective function can potentially satisfy the interests of all the parties.

- Minimizing the *average delay* which includes the lateness and earliness.

$$\sum_{i=1}^n (L_i \max\{(SLT_i - TLT_i - Al_i^+), 0\} + E_i \max\{(TLT_i - SLT_i - Al_i^-), 0\}). \quad (1)$$

- Maximizing the *runway throughput* (or runway utilization). Equivalently, the average of the landing times can be minimized rather than maximizing the number of aircraft landing on the runway (throughput or lead time).

$$\frac{1}{n} \sum_{i=1}^n SLT_i. \quad (2)$$

- Minimizing the *fuel burn cost* (and hence minimizing carbon dioxide emissions). Fuel cost is almost 50% of the operating cost. The fuel burn depends on different factors such as pilot flying techniques, altitude, air speed, aircraft model, aircraft weight (including passengers and cargo), and fuel in the tank. Consequently, the extra fuel burn cost caused by lateness and earliness has to be considered.

$$\sum_{i=1}^n (FB_i^+ \max\{(SLT_i - TLT_i), 0\} + FB_i^- \max\{(TLT_i - SLT_i), 0\}). \quad (3)$$

Since the ALP may involve the simultaneous optimization of various correlated dependent objectives that are not necessarily aligned, a trade-off among the objectives is required. Therefore, they need to be optimized in the form of a weighted multi-criteria objective function.

C. Constraints

A variety of operational constraints can be imposed for the ALP, the most typical of which are the following.

- *Runway Usage*: Each runway can be used by at most one aircraft at a time so either aircraft i lands before j or vice versa.

$$X_{ij} + X_{ji} = 1 \quad \forall i, j \in A, i \neq j. \quad (4)$$

- *Wake Vortex (WV) Separation*: Aircraft have to observe a separation distance to avoid turbulence caused by preceding aircraft.

$$SLT_i + S_{ij} \leq SLT_j + M(1 - X_{ij}) \quad \forall i, j \in A, i \neq j, M \gg 0. \quad (5)$$

- *Time Constraint*: Based on operational and technical considerations such as limited fuel, airspeed, etc., each

aircraft has a maximum and a minimum allowable airborne time which have to be treated as hard constraints.

$$\text{ELT}_i \leq \text{SLT}_i \leq \text{LLT}_i \quad \forall i \in A. \quad (6)$$

A time slot (or time window) assigned to each landing aircraft which typically starts 5 minutes before TLT_i and ends 10 minutes after TLT_i does not necessarily coincide with the time constraint.

- **Time Shifting:** There is limited flexibility in moving the aircraft's landing time either forward or backward in time relative to its estimated landing time. Time shifting is considered rather than position shifting in re-sequencing the aircraft since it can be dependent on aircraft type.

$$(\text{Est}_i - \text{TS}_i) \leq \text{SLT}_i \leq (\text{Est}_i + \text{TS}_i) \quad \forall i \in A. \quad (7)$$

- **Precedence Constraint:** Airline preferences may dictate that one aircraft should land before another.

$$\text{SLT}_i P_{ij} < \text{SLT}_j \quad \forall i, j \in A, i \neq j. \quad (8)$$

IV. FUTURE RESEARCH DIRECTIONS

Although many research papers on the ALP have been published during the last three decades, the majority have not developed methods that have been implemented. The reason could be because the methods may relax or dismiss hard (critical) operational constraints, have unreasonable algorithm runtime, study a static rather than dynamic environment, ignore the requirements of the various stakeholders, or depend on features of a specific airport. Existing research generally considers some of the common and obvious constraints. This research aims to capture vital operational constraints that have been observed from the daily work of controllers in our model building.

Solution approaches for the model remain to be developed. The solutions must be obtained quickly to be of use to air traffic controllers. Since the problem is complex (it is NP-hard), heuristic methods including local search algorithms may be more appropriate than enumerative methods such as dynamic programming which can be computationally demanding.

ACKNOWLEDGMENT

This work has been co-financed by the European Organisation for the Safety of Air Navigation (EUROCONTROL) under its Research Grant scheme.

The content of the work does not necessarily reflect the official position of EUROCONTROL on the matter.

© 2010, EUROCONTROL and the University of Southampton. All Rights reserved.

REFERENCES

- [1] A. Blumstein, "The landing Capacity of a Runway," *Operations Research*, vol. 7(6), pp. 752–763, 1959.
- [2] S. Capri and M. Ignaccolo, "Genetic Algorithms for Solving the Aircraft-Sequencing Problem: The Introduction of Departures into the Dynamic Model," *Journal of Air Transport Management*, vol. 10(5), pp. 345–351, 2004.
- [3] K. Artiouchine, P. Baptiste, and C. Durr, "Runway Sequencing with Holding Patterns," *European Journal of Operational Research*, vol. 189(3), pp. 1254–1266, 2008.
- [4] R. Dear, "The Dynamic Scheduling of Aircraft in the Near Terminal Area," R76-9, Flight Transportation Laboratory, MIT, USA, Tech. Rep., 1976.
- [5] H. N. Psaraftis, "A Dynamic Programming Approach to the Aircraft Sequencing Problem," R78-4, Flight Transportation Laboratory, MIT, USA, Tech. Rep., 1978.
- [6] L. Bianco, P. Dell'Olmo, and S. Giordani, "Minimizing Total Completion Time Subject to Release Dates and Sequence-Dependent Processing Times," *Annals of Operations Research*, vol. 86, pp. 393–415, 1999.
- [7] A. M. Bayen, C. J. Tomlin, Y. Ye, and J. Zhang, "An Approximation Algorithm for Scheduling Aircraft with Holding Time," in *43rd IEEE Conference on Decision and Control, Atlantis, Paradise Island, Bahamas, December 14-17*, 2004.
- [8] H. Balakrishnan and B. Chandran, "Scheduling Aircraft Landings under Constrained Position Shifting," in *AIAA Guidance, Navigation and Control Conference and Exhibit, Keystone, Colorado, USA, August 21-24*, 2006.
- [9] B. Chandran and H. Balakrishnan, "A Dynamic Programming Algorithm for Robust Runway Scheduling," in *Proceedings of the American Control Conference, New York, NY, USA, July 11-13*, 2007.
- [10] H. Lee and H. Balakrishnan, "Fuel Cost, Delay and Throughput Tradeoffs in Runway Scheduling," in *Proceeding of American Control Conference (ACC 08), Seattle, Washington, USA, June 11-13*, 2008.
- [11] A. R. Brentnall, "Aircraft Arrival Management," Ph.D. dissertation, University of Southampton, UK, 2006.
- [12] A. R. Brentnall and R. C. H. Cheng, "Some Effects of Aircraft Arrival Sequence Algorithms," *Journal of Operational Research Society*, pp. 1–11, 2008.
- [13] J. E. Beasley, M. Krishnamoorthy, Y. M. Sharaiha, and D. Abramson, "Scheduling Aircraft Landings - The Static Case," *Transportation Science*, vol. 34(2), pp. 180–197, 2000.
- [14] J. Abela, D. Abramson, M. Krishnamoorthy, and A. D. Silva, "Computing Optimal Schedules for Landing Aircraft," in *Proceeding of 12th National Conference of the Australian Society for Operations Research, Adelaide, Australia, July 7-9*, pp. 71 – 90, 1993.
- [15] G. Stevens, "An Approach to Scheduling Aircraft Landing Times Using Genetic Algorithms," *Honours Thesis, Department of Computer Science, RMIT University, Melbourne, Australia*, 1995.
- [16] V. Ciesielski and P. Scerri, "An Anytime Algorithm for Scheduling of Aircraft Landing Times Using Genetic Algorithms," *Australian Journal of Intelligent Information Processing Systems*, vol. 4(3/4), pp. 206–213, 1997.
- [17] ———, "Real Time Genetic Scheduling of Aircraft Landing Times," in *D. Fogel (Editor), Proceedings of The 1998 IEEE International Conference on Evolutionary Computation (ICEC'98), Anchorage, Alaska, May 4-9*, 1998.
- [18] V. H. L. Cheng, L. S. Crawford, and P. K. Menon, "Air Traffic Control Using Genetic Search Techniques," in *Proceedings of the IEEE International Conference on Control Applications, Hawaii, HA, USA, August 22-27*, 1999.
- [19] J. E. Beasley, J. Sonander, and P. Havelock, "Scheduling Aircraft Landing at London Heathrow using a Population Heuristic," *Journal of Operational Research Society*, vol. 52, pp. 483–493, 2001.
- [20] J. V. Hansen, "Genetic Search Methods in Air Traffic Control," *Computers & Operations Research*, vol. 3, pp. 445–459, 2004.
- [21] X.-B. Hu and W.-H. Chen, "Genetic Algorithm Based on Receding Horizon Control for Arrival Sequencing and Scheduling," *Engineering Applications of Artificial Intelligence*, vol. 18(5), pp. 633–642, 2005.
- [22] X.-B. Hu and E. D. Paolo, "Binary-Representation-Based Genetic Algorithm for Aircraft Arrival Sequencing and Scheduling," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9(2), pp. 301–310, 2008.
- [23] ———, "An Efficient Genetic Algorithm with Uniform Crossover for Air Traffic Control," *Computer & Operations Research*, vol. 36, pp. 245–259, 2009.
- [24] H. Pinol and J. E. Beasley, "Scatter Search and Bionic Algorithms for the Aircraft Landing Problem," *European Journal of operational Research*, vol. 171, pp. 439–462, 2006.
- [25] M. C. Randall, "Scheduling Aircraft Landings Using Ant Colony Optimisation," in *Proceedings of The IASTED International Conference Artificial Intelligence and Soft Computing, Banff, Canada, July 17-19*, 2002.

Author Index

A			
Ahmad, N.	525	Dougui, N.	185
Alonso, R.	371	Drexler, J.	67
Atkin, J. A. D.	131	Duong, V.	401
Augris, H.	263	E	
		Eckstein, A.	3
B		Everdij, M.	449
Badanik, B.	111	F	
Ball, M.	23, 75, 193, 217, 323	Feigh, K.	263
Bayen, A.	85	Ferguson, J.	101
Beboux, G.	379	Feron, E.	11, 409
Bennell, J. A.	529	Finger, M.	419
Bloem, E. A.	443	Fondacci, R.	513
Blom, H. A.P.	443, 465, 449	Fricke, H.	155, 209, 457, 473
Burgain, P.	409	Friedrich, M.	155, 347
Burke, E. K.	131	Fry, D.	271
Buxi, G. S.	305	Fukuda, Y.	165, 237
C		Fürstenau, N.	347
Calderón-Meza, G.	495	G	
Callaham, M. B.	67	Ganji, M.	193
Castelli, L.	201, 433	Gariel, M.	11
Causse, M.	387, 395	Garrow, L. A.	505, 517
Chaloulos, G.	255	Gelnarová, E.	225
Chauhan, A.	31	Gianazza, D.	379
Chen, X.	31	Glover, C. N.	217
Churchill, A. M.	323	Günther, T.	155, 209
Clarke, J.-P.	409	H	
Constans, S.	513	Hansen, M.	61, 75, 305, 331, 425
Corolli, L.	201	Harback, K.	67
D		Hecker, P.	119
D'Innocenzo, A.	487	Henriques, M. R.	53
Dehais, F.	387, 395	Herodes, M.	225
Del Pozo, I.	37	Hoffman, K.	101
Delahaye, D.	179, 185	Hu, M. H.	19, 401, 501, 509
DeLaurentis, D.	271	Hurter, C.	371, 379
Delgado, L.	339	K	
Di Benedetto, M. D.	487	Kageyama, K.	165
Di Matteo, G.	487	Kallos, G.	297
Diana, T.	279		

Kara, A. Q.	101	Papenfuß, A.	347
Kaufhold, R.	119	Pasquini, A.	465
Kim, A. M.	331	Pastor, J.	387, 395
Kotegawa, T.	271	Péran, P.	395
Krozel, J.	289	Piera, M. A.	37
Krupanský, P.	225	Polishhcuk, V.	289
Kumar, V.	363	Popescu, V.	263
Kushta, J.	297	Potts, C. N.	529
L		Pourtaklo, N. V.	23
Labbé, M.	433	Prats, X.	339
Lesire, C.	147	Puchaty, E.	271
Liu, F. Q.	19	Puechmorel, S.	179, 185
Lovell, D.	85, 193, 323	R	
Lu, W.	419	Rabbani, T.	85
Luken, B. L.	505	Rabuty, C.	179
Lulli, G.	201	Rakas, J.	45
Lygeros, J.	255	Rapine, C.	513
Lymeropoulos, I.	255	Ravizza, S.	131
M		Rehwald, E.	119
Mainini, M. J.	245	Rey, D	513
Martin, S.	67	Rudolph, M.	347
Matas, M.	111, 125	Ruiz, S.	37
Medina, M.	173	Rumler, W.	209
Mejzlík, P.	225	Ryerson, M. S.	425
Mesgarpour, M.	529	S	
Mitchell, J. S. B.	289	Sabatini, U.	395
Möhlenbrink, C.	347	Salaün, E.	11, 409
Mongeau, M.	185	Saporito, N.	379
Mori, R.	139	Schmidt, M.	347
Morlang, F.	347	Scholte, J. J.	465
Mumbower, S.	517	Senoguchi, A.	237
N		Serrurier, M.	371
Nayak, N.	279	Sherry, L.	101, 173, 363, 481, 495
Netjasov, F.	449	Shortle, J.	481
Neužil, T.	225	Sinclair, A. E.	521
Nikoleris, T.	61	Singhose, W.	409
P		Solomos, S.	297
Paakko, A.	289	Srisaeng, P.	95
Palme, R.	229	Stefanik, M.	111, 125
		Svoboda, J.	225

T

- Tafazzoli, A. 315
Tastambekov, K. 179
Thiel, C. 457, 473
Tosic, V. 449
Tsao, S. 67

V

- Vela, A. 409
Vidal, A. 45
Vidosavljevic, A. 449
Violin, A. 433
Vlachou, K. 85
Vogel, M. 457
Vormer, F. 401

W

- Wang, Y. J. 401
Weiβhaar, U. 209
Weitz, L. A. 355
Wojcik, L. Jr 67

Y

- Ye, B. 501
Yoon, Y. 75
Yousefi, A. 315

Z

- Zadeh, A. N. 315
Zhang, Y. 31, 279, 509
Zhang, Y. M. 481
Zúñiga, C.A. 37

ISBN 978-1-4507-1468-6

ISBN 978-1-4507-1468-6



A standard UPC-A barcode representing the ISBN 978-1-4507-1468-6. The barcode is composed of vertical black bars of varying widths on a white background. To the left of the barcode, the numbers '9 781450 714686' are printed vertically.