

LITERATURE SURVEY

DOMAIN: APPLIED DATASCIENCE

TOPIC: PREDICTION OF FLIGHT DELAY USING MACHINE LEARNING

1. STUDY OF FLIGHT DEPARTURE DELAY AND CASUAL FACTOR USING SPATIAL ANALYSIS.

AUTHORS: shaowu cheng ,yapingzh ang, siqi Hao, ruiwei liu, xiao lu and qian lu

ABSTRACT: Analysis of flight delay and causal factors is crucial in maintaining airspace efficiency and safety. However, delay samples are not independent since they always show a certain aggregation pattern. Therefore, this study develops a novel spatial analysis approach to explore the delay and causal factors which is able to take dependence and the possible problem involved including error correlation and variable lag effect of causal factors on delay into account. The study first explores the delay aggregation pattern by measuring and quantifying the spatial dependence of delay. The spatial error model (SEM) and spatial lag model (SLM) are then established to solve the error correlation and the variable lag effect, respectively. Results show that the SEM and SLM achieve better fit than ordinary least square (OLS) regression, which indicates the effectiveness of considering dependence by employing spatial analysis. Moreover, the outcomes suggest that, aside from the well-known weather and flow control factors, delay-reduction strategies also need to pay more attention to reducing the impact of delay at the previous airport.

REFER LINK: <https://www.hindawi.com/journals/jat/2019/3525912/>

2. ANALYSIS OF AIRCRAFT ARRIVAL AND DEPARTURE DELAY CHARACTERISTICS

AUTHORS: Eric R. Mueller and Gano B. Chatterji

ABSTRACT: The increase in delays in the National Airspace System (NAS) has been the subject of several studies in recent years. These reports contain delay statistics over the entire NAS, along with some data specific to individual airports, however, a comprehensive characterization and comparison of the delay distributions is absent. Historical delay data for these airports are summarized. The various causal factors related to aircraft, airline operations, change of procedures and traffic volume are also discussed. Motivated by the desire to improve the accuracy of demand prediction in enroute sectors and at airports through probabilistic delay forecasting, this paper analyzes departure and arrival data for ten major airports in the United States that experience large volumes of traffic and significant delays. To enable such an analysis,

several data fields for every aircraft departing from or arriving at these ten airports in a 21- day period were extracted from the Post Operations Evaluation Tool (POET) database. Distributions that show the probability of a certain delay time for a given aircraft were created. These delay-time probability density functions were modeled using Normal and Poisson distributions with the mean and standard deviations derived from the raw data. The models were then improved by adjusting the mean and standard deviation values via a least squares method designed to minimize the fit error between the raw distribution and the model. It is shown that departure delay is better modeled using a Poisson distribution, while the enroute and arrival delays fit the Normal distribution better. Finally, correlation between the number of departures, number of arrivals and departure delays is examined from a time-series modeling perspective.

REFERLINK:

<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.81.9924&rep=rep1&type=pdf>

3. MODELLING AIRPORT EFFICIENCY WITH DISTRIBUTIONS OF THE INEFFICIENT ERROR TERM: AN APPLICATION OF TIME SERIES DATA FOR AIRCRAFT DEPARTURE DELAY

AUTHOR: Ronald Wesonga, Makerere University, School of Statistics and Planning, P.O. Box 7062, Kampala UGANDA

ABSTRACT: The study employs determinants of the aircraft departure delay to estimate airport efficiency. Two main parameters were applied to fit the stochastic frontier model using transcendental logarithmic function where both frontier and inefficiency models were generated. The estimated airport efficiencies over a period of 1827 days applying the half-normal and exponential distributions for the inefficiency error terms were (0.7498; $\sigma^2=0.1417$, $n=1827$) and (0.8181; $\sigma^2=0.1224$, $n=1827$) respectively. The correlation coefficient for the efficiency estimates ($\sigma^2=0.9791$, $n=1827$, $p<0.05$) between the half-normal and exponential distributions showed no significant statistical difference.

Further analysis showed that airport inefficiency was significantly associated with higher number of persons on board, lower visibility level, lower air pressure tendency, higher wind speed and a higher proportion of arrival aircraft delays. The study offers a contribution towards assessing the dynamics for the distribution of inefficient error term to estimate airport efficiency by employing both meteorological and aviation parameters. The study recommends that although either half-normal or exponential distributions could be used; the exponential distribution for the error term was found more suitable when estimating the efficiency score for the airport.

REFERLINK:

<https://gssrr.org/index.php/JournalOfBasicAndApplied/article/view/1229>

4. FLIGHT DELAY CAUSALITY :MACHINE LEARNING TECHNIQUE IN CONJUNCTION WITH RANDOM PARAMETER STATISTICAL ANALYSIS

AUTHORS: Seyedmirsajad Mokhtarimousav and Armin Mehrabi

ABSTRACT: The consequences of flight delay can significantly impact airports' on-time performance and airline operations, which have a strong positive correlation with passenger satisfaction. Thus, an accurate investigation of the variables that cause delays is of main importance in decision-making processes. Although statistical models have been traditionally used in flight delay analysis, the presence of unobserved heterogeneity in flight data has been less discussed. This study carried out an empirical analysis to investigate the potential unobserved heterogeneity and the impact of significant variables on flight delay using two modeling approaches. First, preliminary insight into potential significant variables was obtained through a random parameter logit model (also known as the mixed logit model). Then, a Support Vector Machines (SVM) model trained by the Artificial Bee Colony (ABC) algorithm, was employed to explore the non-linear relationship between flight delay outcomes and causal factors. The data-driven analysis was conducted using three-month flight arrival data from Miami International Airport (MIA). A variable impact analysis was also conducted considering the black-box characteristic of the SVM and compared to the effects of variables indented through the random parameter logit modeling framework. While a large unobserved heterogeneity was observed, the impacts of various explanatory variables were examined in terms of flight departure performance, geographical specification of the origin airport, day of month and day of week of the flight, cause of delay, and gate information. The comprehensive assessment of the contributing factors proposed in this study provides invaluable insights into flight delay modeling and analysis.

REFER LINK: <https://www.sciencedirect.com/science/article/pii/S2046043022000119>

5. A MULTILAYER PERCEPTRON NEURAL NETWORK WITH SELECTIVE-DATA TRAINING FOR FLIGHT ARRIVAL DELAY PREDICTION

AUTHORS: Hajar Alla , Lahcen Moumoun and Youssef Balouk

ABSTRACT: Flight delay is the most common preoccupation of aviation stakeholders around the world. Airlines, which suffer from a monetary and customer loyalty loss, are the most affected. Various studies have attempted to analyze and solve flight delays using machine learning algorithms. This research aims to predict flights' arrival delay using Artificial Neural Network (ANN). We applied a MultiLayer Perceptron (MLP) to train and test our data. Two approaches have been adopted in our work. In the first one, we used historical flight data extracted from Bureau of Transportation Statistics (BTS). The second approach improves the efficiency of the model by applying selective-data training. It consists of selecting only most relevant instances from the training dataset which are delayed flights. According to BTS, a flight whose difference between scheduled and actual arrival times is 15 minutes or greater is

considered delayed. Departure delays and flight distance proved to be very contributive to flight delays. An adjusted and optimized hyperparameters using grid search technique helped us choose the right architecture of the network and have a better accuracy and less error than the existing literature. The results of both traditional and selective training were compared. The efficiency and time complexity of the second method are compared against those of the traditional training procedure. The neural network MLP was able to predict flight arrival delay with a coefficient of determination of 0.9048, and the selective procedure achieved a time saving and a better score of 0.9560. To enhance the reliability of the proposed method, the performance of the MLP was compared with that of Gradient Boosting (GB) and Decision Trees (DT). The result is that the MLP outperformed all existing benchmark methods.

REFER LINK: <https://www.hindawi.com/journals/sp/2021/5558918/>

6. EMPIRICAL ANALYSIS OF AIRPORT NETWORK AND CRITICAL AIRPORTS

AUTHORS: Cong ,Wei, Hu, Minhua, Dong bin, Wang Yanjun , feng ,cheg

ABSTRACT: Air transport network, or airport network, is a complex network involving numerous airports. Effective management of the air transport system requires an in-depth understanding of the roles of airports in the network. Whereas knowledge on air transport network properties has been improved greatly, methods to find critical airports in the network are still lacking. In this paper, we present methods to investigate network properties and to identify critical airports in the network. A novel network model is proposed with airports as nodes and the correlations between traffic flow of airports as edges. Spectral clustering algorithm is developed to classify airports. Spatial distribution characteristics and infraclass correlation of different categories of airports are carefully analyzed. The analyses based on the fluctuation trend of distance-correlation and power spectrum of time series are performed to examine the self-organized criticality of the network. The results indicate that there is one category of airports which dominates the self-organized critical state of the network. Six airports in this category are found to be the most important ones in the Chinese air transport network. The flights delay occurred in these six airports can propagate to the other airports, having huge impact on the operation characteristics of the entire network. The methods proposed here taking traffic dynamics into account are capable of identifying critical airports in the whole air transport network.

REFER LINK: <https://www.sciencedirect.com/science/article/pii/S1000936116000376>

7. ESTIMATING THE FLIGHT DEPARTURE DELAY DISTRIBUTIONS – A STATISTICAL APPROACH WITH LONG- TERM TREND AND SHORT TERM PATTERN

AUTHORS: Yufeng Tu ,Michel O Ball & Wolfgang S Jank

ABSTRACT: In this article we develop a model for estimating flight departure delay distributions required by air traffic congestion prediction models. We identify and study major factors that influence flight departure delays, and develop a strategic departure delay prediction model. This model employs nonparametric methods for daily and seasonal trends. In addition, the model uses a mixture distribution to estimate the residual errors. To overcome problems with local optima in the mixture distribution, we develop a global optimization version of the expectation–maximization algorithm, borrowing ideas from genetic algorithms. The model demonstrates reasonable goodness of fit, robustness to the choice of the model parameters, and good predictive capabilities. We use flight data from United Airlines and Denver International Airport from the years 2000/2001 to train and validate our model.

REFERLINK: <https://www.tandfonline.com/doi/abs/10.1198/016214507000000257>

8. MODELLING CONGESTION PROPAGATION IN MULTISTAGE SCHEDULE WITHIN AN AIRPORT NETWORK

AUTHORS: Xiaoxu Dai, Minghua Hu, Wen Tian, Hao Liu

ABSTRACT: In order to alleviate flight delay it is important to understand how air traffic congestion evolves or propagates. In this context, this paper focusses on the aggravation of airport congestion by the accumulation of delayed departure flights. We start by applying a heterogeneous network model that takes congestion connection/degree into consideration to predict departure congestion clusters. This is on the basis of the fact that, from a micro perspective, the connection between congestion and discrete clusters can be embodied in models. However, the results show prediction to be of high accuracy and time consuming due to the complexities in capturing the connection in congested flights. The problem of being highly time consuming is resolved in this paper by improving the models by stages. Stage partitioning based on the variation of delay clusters is similar to the typical infectious cycle. For heterogeneous

networks the model can describe the congestion propagation and its causes at the different stages of operation. If the connection between flights is homogeneous, the model can describe a more indicative process or trend of congestion propagation. In particular, for single source congestion, the simplified multistage models enable short-term prediction to be fast. Furthermore, for the controllers, the accuracy of prediction using simplified models can be acceptable and the speed on the prediction is significantly increased. The simplified models can help controllers to understand congestion propagation characteristics at different stages of operation, make a fast and short-term prediction of congestion clusters, and facilitate the formulation of traffic control strategies.

REFERLINK:

https://www.researchgate.net/publication/327056786_Modeling_Congestion_Propagation_in_Multistage_Schedule_within_an_Airport_Network

9. EVALUATION AND ANALYSIS OF THE IMPACT OF AIRPORT DELAY

AUTHORS: Fen Zhou, Guosong Jiang, Zhengwu Lu, and Qingdong Wang

ABSTRACT: To characterize the real dynamic process of delay propagation between airports and to understand the mechanism of delay propagation from a global perspective, this paper establishes a network of delay propagation relationships between airports based on causal analysis and conducts an example analysis. First, the flight operating status data processing is used to obtain the arrival delay time series of each airport. Then, the delay propagation relationship between airports is analyzed in pairs through the causal analysis method to obtain the delay propagation relationship network between all airports. Finally, the complex network theory and related indicators are used. The network is analyzed, and these analysis results can provide theoretical support for the formulation of delay propagation mitigation measures.

REFERLINK: <https://www.hindawi.com/journals/sp/2022/7102267/>

10. MODELLING FLIGHT DELAY PROPAGATION IN AIRPORT AND AIRSPACE NETWORK

AUTHORS: Qinggang Wu, Minghua Hu, Xiaozhen Ma, Yanjun Wang, Wei Cong, Daniel Delahaye

ABSTRACT: An Airport-Sector Network Delays model is developed in this paper for flight delay estimation within air transport network. This model takes both airports and airspace capacities into account by iterating among its three main components: a queuing engine, which treats each airport in the network as a queuing system and is used to compute delays at individual airport, a Link Transmission Model, which computes delays at individual sector and transmits all air delays into ground delays, and a delay propagation algorithm that updates flight itineraries and demand rates at each airport on the basis of the local delays computed by the queuing engine and flow control delays computed by the Link Transmission Model. The model has been implemented to a network consisting of the 21 busiest airports in China and 2962 links that represent to 151 enroute control sectors in mainland China, and its performance is evaluated by comparing with the actual delay data and results of Airport Network Delays model. It is found that the proposed model is well-suited for simulating delays in air transport system where either airports or airspace could be the bottleneck of the system.

REFERLINK: <https://hal-enac.archives-ouvertes.fr/hal-01897108/document>

11. COMPARATIVE ANALYSIS ON PROGATION EFFECTS OF FLIGHT DELAYS :A CASE STUDY OF CHINA AIRLINES

AUTHORS: Weiwei Wu, **Cheng-Lung Wu**, Tao Feng's, Zhang Haoyu

ABSTRACT: This paper aims to capture the interdependency among the sequence of flight delays due to airline operations in airports, weather, and air traffic control conditions. A copula function is used to determine the distribution of delay sequence and examine the propagation effects. Using the actual data sourced from an airline in Asia Pacific region, it is found that flight delays could propagate to downstream airports/airlines, where the strength of delays was decreased, passed on, or increased. Considering the possible effects of increased delays under air traffic control or airline factors, scenarios that adjust flight schedules with additional buffer time were created and analyzed. Results show that, by adding buffer time efficiently, flight schedules can become more reliable.

REFERLINK:

[https://www.researchgate.net/publication/322796669 Comparative Analysis on Propagation Effects of Flight Delays A Case Study of China Airlines](https://www.researchgate.net/publication/322796669_Comparative_Analysis_on_Propagation_Effects_of_Flight_Delays_A_Case_Study_of_China_Airlines)

12. AIR TRAFFIC FLOW MANAGEMENT SLOT ALLOCATION TO MINIMIZE PROPAGATED DELAY AND IMPROVE AIRPORT SLOT ADHERENCE

AUTHORS: Nikola Ivanov^a, Fedja Netjasov, Radosav Jovanović, Stefano Starita, Arne Strauss

ABSTRACT: In Europe, one of the instruments at the Network Manager's (NM) disposal to tackle demand-capacity imbalance is to impose ground, i.e. Air Traffic Flow Management (ATFM), delays to flights. To compensate for anticipated delays and improve on-time performance, Aircraft Operators usually embed a buffer time in their schedules. The current practice for allocating ATFM delays does not take into account if flights have any remaining schedule buffer to absorb ATFM delay and reduce delay propagation to subsequent flights. Furthermore, the policy presently employed is to minimize ATFM delays, an order of magnitude of half a minute per flight, while propagated delays are approximately ten times higher. In this paper, we explore the possibility to control ATFM delay distribution in a way so as to minimize delay propagated to subsequent flights, but also to increase flights' adherence to airport slots at coordinated airports. To this aim, we propose a two-level mixed-integer optimization model to solve en-route demand-capacity imbalance problem and further improve airport slot adherence. The rationales behind the research are drawn from practical experience, while the model proposed is compatible with the one currently being used by the NM, making it easy to implement. We test the model on two real-world case studies and conduct ex post analysis to test the effects of violation of model assumptions on results. The results show that it is possible to use the proposed methodology to lower delay propagated to subsequent flights and at the same time to improve airport slot adherence. In addition, they suggest that the current regulatory settings aiming to minimize ATFM delay minutes, as well as operational implementation thereof, are neither necessarily fully aligned with the desires and operating goals of Aircraft Operators, nor they improve the predictability of operations in the network.

REFERLINK: <https://www.sciencedirect.com/science/article/pii/S0965856416300052#!>

13. COMPETING NETWORKS, SPATIAL AND INDUSTRIAL CONCENTRATION IN THE US AIRLINE INDUSTRY

AUTHOR: Aisling J. Reynolds-Feighen

ABSTRACT: The paper uses Gini decomposition analysis to evaluate changes in the spatial distribution and industry shares of total US air traffic, as well as analysing the decomposition components for individual airlines and airports for the period 1990–2002. The paper develops explicit relationships between two of the main decomposition schemes used in the income inequality literature and shows the insights that such analysis may provide for evaluation and examination of air transport networks and traffic distributions. A multi-dimensional Gini and its

decomposition are derived using an adjustment method derived from the relationship between the two Gini decomposition schemes.

REFERLINK: <https://www.tandfonline.com/doi/full/10.1080/17421770701549779>

14. FLIGHT DELAY FORCASTING ANG ANALYSIS OF DIRECT AND INDIRECT FACTORS

AUTHORS: Fujun Wang,Jun Bi ,Dongfan Xie,Xiaomei Zhao

ABSTRACT: The accurate prediction of flight delays is of great significance to airports, airlines and passengers. This paper presents a causal flight delay prediction model developed for a single airport. A long short-term memory network of delay prediction with an attention mechanism (LSTM-AM) is established to predict flight delays and analyse their primary causes. In this model, the direct and indirect factors related to delays are comprehensively considered. LSTM-AM can focus on input data combined with the attention vector to capture the critical time points, which can make the prediction more accurate. The model's performance is verified by actual operational data of Beijing International Airport, one of the busiest airports in the world. Experimental results show that LSTM-AM has better prediction accuracy than baseline algorithms such as some machine learning methods and deep learning methods. The mean absolute error of LSTM-AM is about 8.15 min on the test dataset. The study found that using the predicted results of this paper to release delayed information in advance can effectively alleviate the nervousness of passengers. The critical time point captured by LSTM-AM combined with runway and apron flow control can reduce or eliminate delays of one flight.

REFERLINK: <https://ietresearch.onlinelibrary.wiley.com/doi/full/10.1049/itr2.12183>

