

Predictability Impacts of Airport Surface Automation

Yi Liu*

Department of Civil and Environmental Engineering
107 McLaughlin Hall
University of California, Berkeley
Berkeley, USA 94720
1-510-501-4100
liuyisha@berkeley.edu

Mark Hansen

Department of Civil and Environmental Engineering
114 McLaughlin Hall
University of California, Berkeley
Berkeley, CA 94720

Gautam Gupta, Waqar Malik

University of California, Santa Cruz
NASA Ames Research Center
Moffett Field, CA 94035

Yoon Jung

NASA Ames Research Center
Moffett Field, CA 94035

NOTICE: this is the author's version of a work that was accepted for publication in Transportation Research Part C. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Transportation Research Part C, Vol. 44, pp. 128-145, (1 July 2014)] DOI# [10.1016/j.trc.2014.03.010](https://doi.org/10.1016/j.trc.2014.03.010)

Abstract

Performance analysis for airport surface automation typically focuses on capacity utilization and fuel efficiency. Predictability, while recognized as an important day-of-operation performance goal, has received little attention. One reason could be that applicable predictability metrics have not been developed in the context of airport surface operations management. This study addresses this gap by proposing metrics for predictability performance evaluation. Using these metrics, we assess the predictability impacts of automation on airport surface operations using data from a Spot and Runway Departure Advisor (SARDA) Human-In-The-Loop simulation conducted at NASA's Ames Future Flight Central. We find that SARDA substantially improves many aspects of airfield operation predictability.

Keywords:

Airport surface operations; Gate holding; Automation; Predictability; Performance Metrics

1. Introduction

The Next Generation Air Transportation System (NextGen) is being enabled by a shift to smarter, satellite-based and digital technologies that, combined with new procedures, promise to make air travel more convenient, predictable and environmentally friendly (Medina, et al., 2013; Federal Aviation Administration, 2013). Improving airport surface operations is recognized as a key element of NextGen because airfield resources are intensively utilized and ground operations play a fundamental role in implementing gate-to-gate trajectory management technologies. Recognizing the importance of airport surface operations, considerable research effort has been devoted to modernizing airfield operations under NextGen. One main cause of inefficiency in this domain is that, under high traffic conditions, multiple aircraft pushback at around the same time and contest for the runway. This leads to many aircraft taxiing to the runway simultaneously and long runway queues as well as congestion effects on taxiways. One means of addressing this problem that has received much attention in the literature is departure metering.

Multiple technologies have been proposed for departure metering (Malik, et al., 2010; Brinton et al., 2011; Nakahara et al., 2011; Simaiakis et al., 2011; Simaiakis, 2013). One approach to departure metering is called N-control, which maintains efficient runway utilization while controlling the queue length by metering pushbacks from the gate (Simaiakis et al., 2011; Simaiakis, 2013). In this technique, the suggested pushback rates are only provided to ground controllers whose primary responsibility is to maintain separation and a smooth flow of aircraft on taxiways. Another method, called Collaborative Departure Queue Management (CDQM), manages the length of the runway queue by assigning flight operators taxiway entry slots according to ration-by-schedule principle (Brinton et al., 2011). A method similar to CDQM was developed and implemented at the John F. Kennedy International Airport in New York, and is currently under use (Nakahara et al., 2011). Both the N-control and CDQM methods manage runway queue length by controlling operations in the ramp only. Another surface traffic management system, known as the Spot And Runway Departure Advisor (SARDA), extends auto assistance to other areas: taxiway, queue area and runway. Specifically, SARDA provides advisories on actual pushback time, sequence and timing for spot release, sequence for take-offs, and sequence for active runway crossings (Jung et al., 2010; Malik et al., 2010; Gupta et al., 2012; Hoang et al., 2011). Moreover, the SARDA advisories are provided to both ground and local controllers. The local controller is responsible for safe and efficient runway operations, including take-off, landing and runway crossings.

In the existing literature on performance evaluation of these automation techniques, attention has been focused on throughput increases, delay reductions, and fuel savings (Simaiakis et al., 2011; Nakahara and Reynolds, 2012; Gupta, et al., 2013). Predictability, while recognized as an important performance goal by various stakeholders (ATSPFG 1999; Bradford, et al., 2000), has received little or no attention. In the context of airport surface operations, high predictability allows a multitude of direct benefits, such as reductions in communication and controller workload and better adherence to 4-dimensional (time-space) trajectories. Besides the operational benefits, high predictability may also deliver direct financial and environmental benefits by enabling greater use of single engine taxiing. If pilots know the departure clearance time in advance with

certainty, they can use a single engine while taxiing until the time when they should start the second engine for take-off. With less predictability about departure clearance time, they must start the second engine earlier than necessary to make sure they are not caught short when cleared for departure, which consumes more fuel and generates more emissions. Given the importance of predictability and its potential benefits, it is desirable to understand how automation technologies affect operational predictability performance. However, the lack of predictability metrics hampers our ability to assess the predictability impact of automation tools.

To fill this gap and provide a new perspective on evaluating the impacts of automation technologies on airport surface operational performance, this study seeks to define and quantify predictability in the context of airport surface operation management. Using SARDA simulation data, we present a comprehensive assessment of the predictability impacts of airport surface automation. A wide range of such impacts is considered, which includes variability in taxi-out time, predictability of take-off time and take-off sequence, entropy of the airfield state, and perceived predictability from users. With the necessary data, the proposed predictability analysis could be repeated for other surface traffic automation tools and for other airports.

While this paper is mainly a scientific study based on the results of an automation tool simulation experiment, it is also an exploration of the meaning of “predictability” in the airport surface context, and more broadly in Air Traffic Management (ATM). There is wide consensus in the aviation community that predictability is an important performance goal for ATM (Bradford, et al., 2000; Liu and Hansen, 2014). International Civil Aviation Organization (ICAO, 2005) emphasizes that predictability is essential to airspace users as they develop and operate their schedules. Service providers such as the Federal Aviation Administration (FAA) identify predictability as a key performance goal (Knorr et al., 2000; FAA, 2011). Anecdotally, flight operators respond very positively to the idea of improving predictability. However, the definition of predictability is elusive to the community. Thus, a contribution of this paper is to elucidate what predictability means, as well as how it might be measured, in the context of airport surface management. Moreover, through such efforts, claims of predictability improvements can become clear statements subject to rigorous analysis, rather than vague and unverifiable sentiments.

The remainder of the paper is organized as follows. In Section 2, the current related literature on predictability is summarized. In Section 3, the SARDA system and the experimental setup of the Human-in-the-Loop simulation are introduced, and the methods that are used to analyze SARDA predictability impacts are briefly discussed. In Sections 4 to 6, predictability performance is assessed and compared for the baseline case—without SARDA—and the advisory case—with SARDA—from multiple perspectives. Finally, the paper is concluded in Section 7.

2. Related Literature on Predictability

The majority of the literature on predictability in transportation assesses predictability by measuring variability in the ‘travel time’, which could be a road trip travel time, gate-to-gate time of a given flight, or taxi-out time of an aircraft on the airfield. There is a variety of variability measurements: difference between actual trip time and scheduled trip time (Kho et al., 2005), standard deviation of travel time distribution (Bates et al.,

2001; Lomax et al., 2003; Ettema and Timmermans, 2006; Riikka and Paavilainen 2010), standard deviation over the mean travel time (Taylor 1982; Lomax et al., 2003), difference between travel time percentiles (Ettema and Timmermans, 2006; Gulding et al., 2009; Li and Rose, 2011) and difference in expected and actual travel delays (Cohen and Southworth, 1999; Liu and Hansen, in press). In the context of advanced traveler information system, a few studies have examined the impact of intelligent transportation system on trip reliability, where similar predictability measures are used: standard deviation in travel time (Levinson, 2003) and difference between actual trip time and scheduled trip time (Kristof et al., 2005). None of these studies explicitly consider the temporal aspect of predictability. In contrast, Ball et al. (2000) find that error in predicting flight departure time decreases as the departure of a flight approaches. They proposed a metric termed integrated predictive error that takes this effect into account.

Other studies are not concerned with predictability per se, but rather focus on methods for predicting travel time on the basis of the information available prior to the commencement of the travel. Many of these studies focus on road networks and are motivated by the increasing use of routing and navigation decision support tools (Borokhov et al., 2011). Linear regression, based on a combination of the current information—system variables—and historical travel time information, has served as one of the main methodologies (Kwon et al., 2000; Zhang and Rice, 2003; Rice and Zwet, 2004). Considering the importance of prediction timeliness in the application, the algorithms are usually designed to be simple, fast and scalable (Rice and Zwet, 2004). In these studies, the travel time predictions are modeled to guide travel decisions but are not linked to performance measurement.

In the broader literature on systems, a concept closely related to predictability is entropy—a measure of the uncertainty in a random variable. Entropy, since its introduction into information theory by Shannon (1948), has been used to measure unpredictability of a set of possible events. Entropy has also been used to characterize stochastic processes, defined as an indexed sequence of random variables that can take values from a set of possible states. Studies have been conducted to validate the application of entropy analysis in stochastic processes (Cover and Thomas, 1991; Ciuperca and Girardin, 2005; Jacquet et al., 2008). The entropy rate is defined for all stationary processes but it is more widely linked to Markov Chain (MC) processes, in which the memoryless property leads to easy application (Cover and Thomas, 1991).

3. Methodology

3.1 Spot And Runway Departure Advisor

SARDA, developed at NASA Ames Research Center, uses time-based metering of aircraft to reduce the number of aircraft on the airport surface. SARDA's domain of interest covers the airport surface where departure and arrival aircraft operate, including ramps, taxiways, and runways. These areas on the airport surface are under the control of different stakeholders, such as flight operators, airport operators, and the Federal Aviation Administration (FAA). The SARDA concept assumes a collaborative framework for obtaining flight movement and related operations information from these stakeholders, and providing flight specific metering advisories back to them. For each aircraft, SARDA provides the metering advisories at three main locations: the gate, spot and runway. It provides ramp controllers (flight operators) with gate pushback times,

ground controller (FAA) with spot release times and local controller (FAA) with sequence advisories for runway (take-off or runway crossings). By tactically controlling the release of aircraft from the gates and the spots, SARDA effectively shifts the delays from the taxiways and runways to the gates. The delays incurred at the gates are with the aircraft engines off, and provides benefits to the airlines in fuel savings and potentially more time for connecting passengers to make their flights. Moreover, metering of aircraft at the gates reduces the number of aircraft on the movement area at any time, and this can potentially have benefits in increased predictability.

The main algorithm for generating advisories in SARDA is based on the spot release planner (SRP), a two-stage algorithm for providing metering advisories (Malik et al., 2010; Hoang et al., 2011; Gupta et al., 2012). The first stage is a runway scheduler that provides optimal sequence and times for runway usage (take-off times for departures and runway crossing times for arrivals). The generated runway schedule complies with various constraints (wake vortex separation, miles-in-trail, Traffic Management Initiatives (TMI), and others) and is optimal for system delay (total delay for all aircraft). The second stage of the SRP determines times to release aircraft from gates or assigned spots to meet the optimal departure schedules. Uncertainties in the system, such as, taxi speed, pilot response to controller taxi clearances, and interaction among taxiing aircraft, is mitigated by executing the algorithm periodically to generate updated optimized solutions.

3.2 Human-in-the-Loop Simulation

In May 2012, the SARDA concept was evaluated in a human-in-the-loop simulation at the Future Flight Central (FFC) facility at the NASA Ames Research Center. The FFC facility can simulate airport surface operations in real-time and provides a 360-degree, full-scale, computer-generated out-the-window view. At FFC, two software components were primarily used for implementation of SARDA: the Airspace Traffic Generator (ATG) and the Surface Management System (SMS). The ATG system was used to generate tracks for aircraft on the surface or in the airspace near the airport. ATG fed the flight plan and track information to SMS, which displayed the flight on its “radar” map. SMS also communicated with the optimization algorithm (scheduler) and obtained the metering advisories, which were then shown to the controllers through an Electronic Flight Strip (EFS) display system. The controllers relayed the advisories over voice radio to the pseudo-pilots. The pseudo-pilots adjusted the aircraft movement through inputs into ATG, and the updated aircraft states were transmitted back to SMS. ATG updated the track information every second and the advisories to the controllers were updated every 10 seconds.

The airport area was modeled after the Dallas-Fort Worth International Airport (DFW) and daytime traffic for East side in the South-flow configuration was simulated assuming good weather condition, which enables operations under visual flight rules. In the South-flow configuration shown in Figure 1, runway 17R is used for departures and 17C for arrivals. Due to staffing limitations, arrivals on 17L and 13L and bridge traffic from West side of airport were not simulated. Arriving aircraft need to cross the departure runway before taxiing in to the gates. The taxiway has two lanes and there are three lanes in the queue area. For departure aircraft, there are three nominal taxi paths defined from taxiway entry to runway entry: outer, inner and full routes. As shown in Figure 1, each nominal trajectory corresponds to a lane in the queue area.

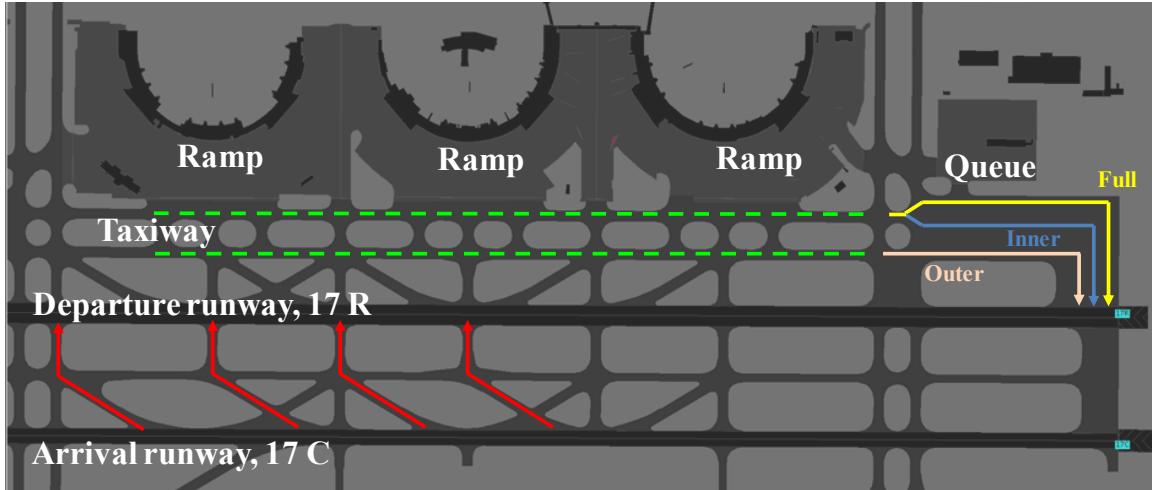


Figure 1. Configuration Map on the East Side of DFW Airport

Four traffic scenarios were generated based on January 2012 DFW traffic; two of these were called medium traffic (1.2x current traffic) and two were called heavy traffic (1.5x current traffic). The medium scenarios had 40 departures over 50 minutes, while heavy scenarios had 50 departures in that span. In each scenario, a few departing aircraft (around 5) were subject to a TMI. To comply with the TMI, they needed to depart at a specified time. These aircraft were provided with their desired take-off time along with a compliance window of one minute.

Six recently retired controllers participated in the experiments over a period of three weeks, with each pair running scenarios for an entire week. During each week, a pair of controllers took turns as ground or local controller and was asked to control traffic under two conditions: advisory and baseline. Under advisory conditions, the SARDA-calculated advisories were shown to the controllers, who were asked to adhere to them as closely as possible. The departure pushback advisories were communicated to a ramp automation agent that controlled the pushback from the gates. In the baseline case the scheduler was not run and the controllers were asked to use their experience in controlling the aircraft. In this scenario, the ramp automation agent pushed aircraft from the gates at their scheduled times. There were a total of 48 data-collection runs in the three weeks, with six runs of each scenario in either baseline or advisory mode.

3.3 Predictability Analysis

Results from the May 2012 simulation have been analyzed to assess the impacts of SARDA on taxi-times, delays, throughput, emissions, and human controllers in several other papers (Gupta et al., 2013; Hayashi et al. 2013). We focus here only on impacts related to predictability.

The SARDA experiments yield detailed operational data—the location of each active aircraft in every second. In addition, controller surveys were conducted. Our predictability analysis is based on the operational data and the answers to the survey questions that were included to gauge controller perspectives on issues related to predictability.

Based on the literature review and our qualitative understanding of the impacts of SARDA, the predictability analysis focused on five main aspects. These included:

- Taxi-out time variability. As noted in the literature review, much of the prior work on transport predictability focuses on trip time variability. In the context of airport surface operation, the “trips” of greatest interest are from the gate to the runway. As noted above, reducing this variability is important for several reasons, including the ability to taxi with a single engine and to better predict when the aircraft will arrive at the destination. Hao and Hansen (2013) find that taxi-out time is the largest source of variability in total block time—the time from leaving the origin gate to arriving at the destination gate.
- Predictability of take-off time. Regression models were constructed to explain and forecast flight take-off time based on variables with values known at the time of actual pushback, and the predictive ability of these models was assessed. The contribution of differences between controllers to taxi-out time variation was also studied using statistical models.
- Sequence predictability. Departing aircraft enter the queue area in some sequence, and depart in some sequence. Moreover, one can calculate a scheduled departure sequence based on scheduled pushback time and unimpeded taxi time from a given gate. As used here, sequence predictability refers to how closely the departure sequence matches the queue entry or scheduled departure sequence. Departing pilots use their place in the departure sequence for a variety of tactical decisions. If sequence predictability is higher, then they will have a better idea of their place in the sequence earlier. Likewise, if the departure sequence could be closely predicted from the scheduled departure and taxi-out times, this would yield predictability benefits.
- Entropy analysis. As discussed above, entropy is a measure of unpredictability. Thus metrics based on entropy are natural candidates for predictability analysis. There are well-established measures of entropy for MC processes in steady-state. Different components of the airfield can be viewed as such systems, allowing these metrics to be applied.
- Controller surveys analysis. Controllers were asked questions at the end of each simulation run (post-run survey) and at the end of their participation in the experiment (post-participation survey). Questions were intended to elucidate opinions related to predictability.

With the exception of the controller post-participation surveys, our analysis of predictability impacts of SARDA is based on comparisons of results obtained from performing the identical analysis on the operational and survey data generated from the baseline and advisory cases. Since the experimental design ensures that the sets of the simulation runs of the two cases have identical taxi-out schedules, pairs of controllers, and weather conditions, the differences between the results obtained can be attributed to the impact of SARDA. In the case of the controller post-participation survey, this approach was not possible, since controllers answered the survey after the completion of all their simulation runs. The survey questions asked the controllers subjective views on differences between the SARDA and baseline systems that pertain to predictability. Findings from this survey are therefore less scientifically rigorous than those of the other parts of our study.

4. Taxi-out Time Predictability

4.1 Time-dependent Variability in Taxi-out Time

Higher predictability in taxi-out time is equivalent to better prediction of the roll time, which will benefit operations on the airfield and improve adherence to 4-D trajectories. In the vast majority of existing literature, predictability is defined by measuring the variation of time spent in different phases of flight (Knorr et al., 2000; Eurocontrol, 2012). In this context, higher predictability results from less variability in taxi-out time. While such ‘gross’ variability is important, a more refined analysis should also consider how this variability changes as the flight progresses through various stages from gate pushback to runway take-off.

To see this, consider two airports, both of which have the same variability (standard deviation) of taxi-out time. However, in the case of Airport A, the taxi-out time is known with certainty a short time after pushback, while in the case of Airport B, the uncertainty persists until later in the departure process. Following the definition in Knorr et al. (2000) and Eurocontrol (2012), the level of predictability in taxi-out time for the two flights would be considered similar. However, because in the case of Airport A, precise taxi-out time is available at an earlier stage, predictability is greater. Cases like this motivate the consideration of timeliness of information in the predictability evaluation.

Generally, uncertainty in remaining taxi-out time decreases through the various stages of the taxi-out process. Measures of predictability should reflect this by taking into account the timeliness as well as the accuracy of information. One way to do this is to integrate uncertainty in remaining taxi-out over time, beginning at scheduled pushback. The resulting measure of unpredictability, α_{up} , is:

$$\alpha_{up} = \int_{t_s}^{t_w} \sigma(t) \cdot dt \quad (1)$$

where, t_s is scheduled pushback time; t_w is actual roll time; and $\sigma(t)$ is standard deviation of the remaining taxi-out time at time t . According to this formula, unpredictability is measured as the integral of standard deviation of remaining taxi-out time along the taxi-out process. In principle, $\sigma(t)$ can be evaluated at any time. However, it is more practical to evaluate this metric by considering the standard deviation at certain discrete “milestones” of the taxi-out process. In this analysis, standard deviation is calculated at six such events: scheduled pushback, actual pushback, taxiway entry, queue entry, runway entry, and takeoff roll. “Takeoff roll” is the event when the aircraft is lined with the runway centerline and is moving forward with the intent to takeoff. With this, the unpredictability equation becomes:

$$\alpha_{up} = \sum_{i=1}^{i=5} (\bar{t}_{i+1} - \bar{t}_i) \cdot \sigma(t_i) \quad (2)$$

where, i is the index for the event—1 for scheduled pushback, 2 for actual pushback, 3 for taxiway entry, 4 for queue entry, 5 for runway entry, and 6 for actual roll time. This metric is calculated for four scenarios defined by traffic level (high and medium) and use of SARDA (advisory and baseline). For each scenario, \bar{t}_i is the average time after scheduled pushback when event i occurs, across all the aircraft in the simulation runs for that scenario. $\sigma(t_i)$ is the standard deviation of the remaining taxi-out time at event i for

all the aircraft. Unpredictability based on this formula reflects the average level of unpredictability for all the flights rather than for each individual flight. If take-off time were a known constant at scheduled pushback time, then unpredictability would be zero.

Changes in standard deviation of remaining taxi-out time over time are shown in Figure 2 for baseline and advisory cases at both traffic levels. We will explain the results using the top-right plot—the scenario with high traffic and SARDA advisories—as an illustration. The horizontal axis represents average time that has passed after scheduled pushback for all the aircraft in the scenario. Standard deviation in total taxi-out time at scheduled pushback, $i = 1$, is about 5 min. Average gate-holding, i.e., the average time difference between scheduled pushback and actual pushback, is 4.4 min. When aircraft actually pushed back from the gate, $i = 2$, standard deviation drops to about 2.3 min which is less than half of before. When aircraft enter the queue area, $i = 4$, standard deviation in the remaining taxi-out time is around 1 min. Average time that aircraft spend in the queue area is 3.3 min. Then aircraft taxi to the runway and take off, and the total taxi out time is 14.8 min on average. For this scenario, unpredictability is 39.0 min^2 which is the area below the blue step plot.

At the same high traffic level, the baseline scenario turns out to have much larger unpredictability, as shown in the plot on the top-left. This happens for two reasons: more variation in total taxi-out time and slower reduction as the taxi-out proceeds. The standard deviation does not go down at the actual pushback because of the absence of gate-holding. Moreover, uncertainty in taxi-out time is high until aircraft taxi into the runway, $i = 5$, where there is—not surprisingly—a big drop in the standard deviation. As a result, the area below the blue step plot is much larger than in the advisory case, indicating less predictability. Percent reductions in this metric under SARDA are 46% under high traffic and 39% under medium traffic.

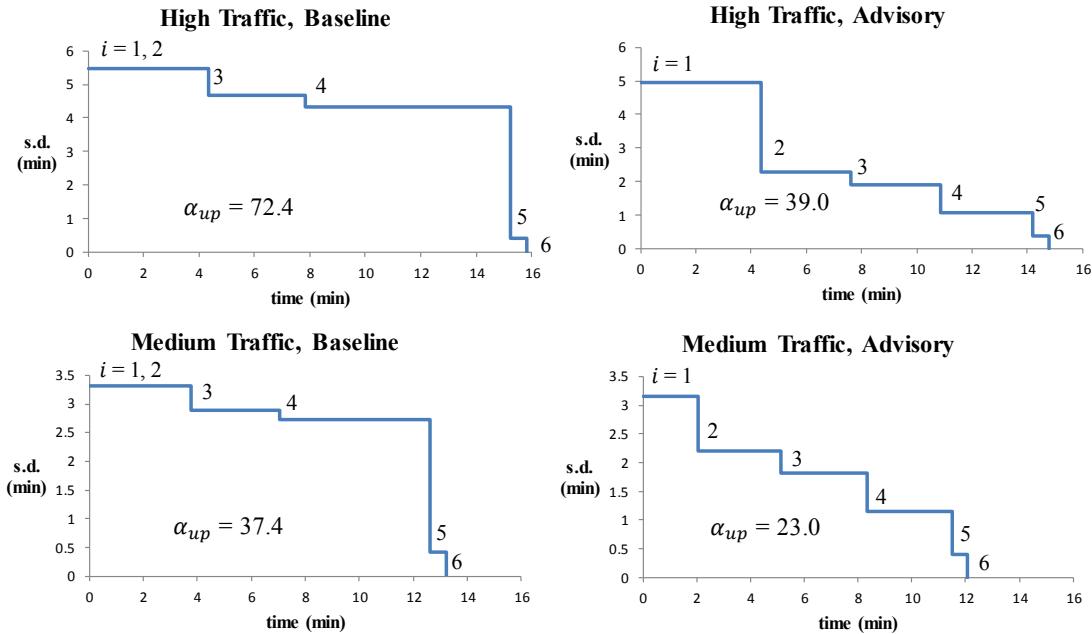


Figure 2. Unpredictability and Time-dependent Variances in Taxi-out Time, Baseline and Advisory, High and Medium Traffic Levels

In summary, uncertainty in roll time declines faster with advisories at both high and medium traffic levels. This is largely due to gate holding under the advisory case, which enables a large proportion of the variability in taxi-out time, especially in the case of high traffic, to be absorbed at the gate before actual pushback. Moreover, even after the aircraft pushback from the gate, uncertainty in the remaining taxi-out time decreases at a somewhat faster rate for the advisory runs. By the time that aircraft enter the queue area, the vast majority of the uncertainty in taxi-out time has been absorbed in the advisory runs whereas in the baseline runs almost all of the uncertainty persists through this stage.

4.2 Model-based Taxi-out Time Predictability

In Section 4.1, we measured predictability based on the variability in taxi-out time and how the variability in remaining taxi-out time diminishes over the different stages of the departure process. In this section, we also focus on taxi-out time, but instead of considering variability per se, we consider the ability to explain and predict taxi-out time, given the information available at pushback. To this end, we estimate and assess the predictive performance of multiple regression models of taxi-out time—from pushback to the beginning of take-off roll—estimated on explanatory variables that are known at the time of pushback.

One potential source of taxi-out time variability is controllers. One might expect that the use of SARDA advisories lessens the impact of inter-controller differences on taxi-out times. This is explored in Section 4.2.3, by adding controller effects to the regression models.

4.2.1 Regression Analysis of Taxi-out Time

In the literature, taxi-out time regression models usually involve explanatory variables that are not estimable at pushback, such as the number of aircraft that will push back while the reference aircraft is taxiing out (Idris et al. 2002; Legge and Levy, 2008; Chauhan 2010; Clewlow et al., 2010; Zhang et al., 2010). While such models may shed light on factors that affect aircraft taxi-out time, they are not capable of predicting the taxi-out time at pushback.

Since we are interested in models that can be used predictively, our regression models use only variables whose values can be known at pushback. Variables that satisfy this requirement in our data include the taxi distances, and aircraft counts at different airfield locations at the time of pushback. Depending on their planned meter fixes, departures on 17R can take one of the two divergent area navigation (RNAV) headings. In the advisory case, aircraft on the same RNAV heading were assigned the same surface route (shown in Figure 1)—the inner route for one heading and the outer route for the other, whereas aircraft under TMI constraints were assigned to the full route. Moreover, SARDA determined and provided taxi trajectory at pushback. As a result, the taxi trajectory is known with certainty at pushback, which allows us to estimate the total taxi distance with high accuracy and further to decompose that distance into four parts: ramp, taxiway, queue area, and on the runway prior to take-off roll. When advisories were absent, the full route was also reserved for TMI flights. Controllers were asked to assign departure routes to other aircraft based on their departure fixes, similar to the advisory case. Therefore, the taxi-out trajectory can be well approximated at pushback, as can the taxi distances in each area. According to the subject matter experts and controllers, the consistency in departure route assignment is common during a controller shift. On the

other hand, in other situations, a diverse set of nominal taxi trajectories are available for a given gate and runway pair at an airport (Zhang and Wang, 2011), which may generate uncertainty in taxi-out trajectory. In this case, the taxi-out time may be predicted with a confidence interval depending on the variance in taxi distances or a set of taxi-out time may be predicted according to the taxi trajectory. In general, therefore, the taxi-out distances will not be known with certainty at pushback in the baseline case. Thus the analysis here, by assuming that taxi distance is known at pushback both with and without advisories, may underestimate the taxi time predictability improvement from automation.

Another set of variables that we could easily observe are counts of aircraft in each area. Similar to the distance variables, there are four count variables at the actual pushback time, the numbers of aircraft in the ramp area, on the taxiways, in the queue area, and in the arrival area. In the simulation, taxi-in aircraft are simulated until they exit the taxiway. Therefore, count of aircraft on the taxiways includes taxi-in and taxi-out aircraft, whereas the count in the ramp area only considers taxi-out aircraft. Aircraft that are considered in the arrival area are the arrival aircraft that have landed but have not crossed the departure runway (Runway 17R) yet. With this, the general equation for taxi-out time regression models is expressed as:

$$Y^f = \beta_0 + \sum_{i=1}^4 \beta_{D,i} \cdot D_i^f + \sum_{i=1}^4 \beta_{N,i} \cdot N_i^f + \varepsilon^f \quad (3)$$

where, Y^f is the taxi-out time (in seconds) between actual pushback and actual take-off for flight f ; D_i^f is the value of distance variable i (in meters) for flight f , $i = 1, 2, 3, 4$; N_i^f is the value of count variable i for flight f , $i = 1, 2, 3, 4$; $\beta_{D,i}$ is the regression coefficient for distance variable i , $i = 1, 2, 3, 4$; $\beta_{N,i}$ is the regression coefficient for count variable i , $i = 1, 2, 3, 4$; β_0 is the intercept; ε^f is a stochastic error term for flight f .

Descriptive statistics for the explanatory variables are summarized in Table 1 for four scenarios: high traffic with advisory, high traffic without advisory, medium traffic with advisory, and medium traffic without advisory. The number of aircraft in the ramp area is counted as the number of aircraft in the same ramp as the reference aircraft, rather than the total number of aircraft in all the ramp areas. Both departure aircraft that are taxiing out and landed aircraft that are taxiing in are counted for aircraft on the taxiway. At both traffic levels, with advisory, there is less traffic in the ramp area and in the queue area on average, whereas taxiway traffic is slightly higher. The baseline scenarios have greater variability in the count variables, with the exception of the arrival area counts. Statistics for the distance variables are about the same for the four scenarios, reflecting similar distributions of gate assignments.

Table 1. Descriptive Statistics for Explanatory Variables in the Taxi-out Time Regression Models, Baseline and Advisory, High and Medium Traffic Levels

	High Traffic							
	Baseline				Advisory			
	Min	Mean	Max	Std	Min	Mean	Max	Std
# in the ramp	0	2.05	8	1.94	0	1.05	4	0.95
# on the taxiway	0	3.87	12	3.11	0	4.61	11	2.52
# in the queue area	0	4.79	14	4.14	0	2.99	7	1.89

# in the arrival area	0	1.72	5	1.38	0	2.07	6	1.26
Ramp distance (m)	195.3	371.6	826.9	136.6	196.1	375.1	827.4	136.7
Taxiway distance (m)	403.2	1381.3	2357.9	617.9	403.2	1383.7	2357.9	616.7
Queue area distance (m)	478.0	615.0	822.6	116.9	478.6	615.0	823.3	116.9
Runway distance (m)	64.7	138.9	262.9	61.6	65.9	139.00	262.93	61.8
Medium Traffic								
	Baseline				Advisory			
	Min	Mean	Max	Std	Min	Mean	Max	Std
# in the ramp	0	1.51	5	1.23	0	1.07	4	0.98
# on the taxiway	0	3.72	12	2.97	0	4	9	2.51
# in the queue area	0	3.5	12	3.33	0	2.51	6	1.87
# in the arrival area	0	1.54	6	1.17	0	1.62	6	1.12
Ramp distance (m)	195.7	360.5	826.7	139.8	195.7	362.7	827.4	139.4
Taxiway distance (m)	403.2	1349.5	2357.3	595.0	485.2	1359.0	2353.7	593.9
Queue area distance (m)	477.9	617.1	821.9	113.4	478.5	611.7	823.4	111.3
Runway distance (m)	64.7	139.7	263.8	60.4	64.7	137.0	261.5	58.7

We first estimated full models with all the distance and counts variables for the four scenarios. We then removed the terms that are not significant at the 0.05 level in the full models and used the remaining as the final models for further study. Estimation results from the final linear taxi-out time regression models are shown in Table 2.

Table 2. Results on Taxi-out Time (second) Regression, Final Models, Baseline and Advisory, High and Medium Traffic Levels

	High Traffic				Medium Traffic			
	Baseline		Advisory		Baseline		Advisory	
	Estimate	St. err.	Estimate	St. err.	Estimate	St. err.	Estimate	St. err.
Intercept	-82.80	75.49	127.60	15.93***	309.46	34.76***	77.59	19.64**
# in the ramp	45.74	6.17***	5.85	2.67*	32.66	7.15***	—	—
# on the taxiway	35.31	4.17***	5.74	1.11***	20.69	2.98***	9.13	1.32***
# in the queue area	8.11	3.14*	6.22	1.73***	8.51	2.57***	7.72	1.81***
# in the arrival area	—	—	9.65	2.45***	—	—	—	—
Ramp distance	0.89	0.08***	0.37	0.02***	0.38	0.06***	0.33	0.02***
Taxiway distance	0.11	0.02***	0.15	0.004***	0.13	0.02***	0.15	0.01***
Queue area distance	0.43	0.09***	0.13	0.02***	—	—	0.23	0.03***
Runway distance	—	—	—	—	—	—	—	—
Number of observation	535		542		423		418	
Residual standard error	253.2		56.8		164		59.44	
Adjusted R ²	0.41		0.83		0.33		0.80	

Values with *** are significant at 0.1% level; Values with ** are significant at 1% level; Values with * are significant at 5% level.

The first four columns are regression results for high traffic level. Positive coefficients for count variables show, as expected, that taxi-out time will be longer if there are more aircraft at different parts of the airfield. The count coefficients are, however, smaller in the advisory model, indicating smaller incremental effect from each aircraft. This suggests that SARDA mitigates the influence of other airfield traffic on the taxi-out time of a given aircraft. The number of aircraft in the arrival area affects taxi-out time in the SARDA high traffic cases but not the baseline one. This is probably because arrivals are considered in the SARDA algorithm. Coefficients for ramp distance and queue area distance are much larger in the baseline case, indicating that without SARDA “stop-and-go” conditions are more prevalent. At medium traffic level, results are similar. The main difference is that in the advisory case, the numbers of aircraft in the ramp and arrival area do not have significant impact on taxi-out time.

The standard deviation of the future prediction on taxi-out time is expected to be much smaller with advisory in high traffic models, since residual standard error in the advisory model is only about one fifth of that in the baseline model. Moreover, with a larger adjusted R², the model for the advisory case has better explanatory power. Similar results are found for the medium traffic level models. Thus, use of SARDA advisories should result in more accurate model-based taxi-out predictions. To test this hypothesis, inequality analysis is performed to compare the accuracy of prediction from the regression models in 4.2.2.

4.2.2 Inequality Analysis of Taxi-out Time Prediction

In this section, we measure the taxi-out time prediction errors from the previous regression models, and investigate the nature of the errors by studying the error proportions (Theil 1966; Kim and Hansen, 2010). In absolute terms, taxi-out time prediction errors are measured by Root-Mean-Square (RMS) prediction error:

$$RMS = \sqrt{\frac{1}{n} \sum_{i=1}^n (TO_i^P - TO_i^R)^2} \quad (4)$$

where, TO_i^P is predicted taxi-out time; TO_i^R is realized taxi-out time; n is total number of observations.

The inequality coefficient, U , calculated as the ratio of RMS prediction error to the square root of the mean square of the realizations, is used to quantify the prediction error in relative terms:

$$U = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (TO_i^P - TO_i^R)^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (TO_i^R)^2}} = \frac{\sqrt{\sum_{i=1}^n (TO_i^P - TO_i^R)^2}}{\sqrt{\sum_{i=1}^n (TO_i^R)^2}} \quad (5)$$

where, U is zero if and only if the predictions are all perfect; U is valued as one when the prediction procedure leads to the same RMS error as would be obtained if the predicted value were always 0. In other words, by using the inequality coefficient one measures the seriousness of a prediction error by the quadratic loss criterion in such a way that 0 corresponds with perfect prediction and unity with the loss from always predicting 0. The prediction error can be decomposed in three components: error in central tendency, error due to unequal variation, and error due to imperfect covariation. A convenient way to handle such decomposition is to divide each of the three terms by their sum. This leads to:

$$\begin{aligned} U_m &= \frac{(\overline{TO^P} - \overline{TO^R})^2}{\frac{1}{n} \sum_{i=1}^n (TO_i^P - TO_i^R)^2} \\ U_s &= \frac{(S_P - S_R)^2}{\frac{1}{n} \sum_{i=1}^n (TO_i^P - TO_i^R)^2} \\ U_c &= \frac{2(1-r)S_P S_R}{\frac{1}{n} \sum_{i=1}^n (TO_i^P - TO_i^R)^2} \end{aligned} \quad (6)$$

where, $\overline{TO^P}$ and $\overline{TO^R}$ are the means of predicted and realized taxi-out time; S_P and S_R are the uncorrected standard deviations of the predicted and realized taxi-out time; r is the correlation coefficient of predicted and realized values:

$$r = \frac{\frac{1}{n} \sum_{i=1}^n (TO_i^P - \overline{TO^P})(TO_i^R - \overline{TO^R})}{S_P S_R} \quad (7)$$

where, U_m is the bias proportion reflecting error in central tendency; U_s is the variance proportion, valued as zero when the standard deviation of predicted and realized changes are equal; U_c is the covariance proportion, valued as zero when the prediction and realization are perfectly positively linear correlated.

In each week, there are four different simulation runs for each scenario—defined by traffic level and SARDA advisory use. For each scenario, taxi-out time is regressed using one week of simulation runs based on the same model specification as that for the corresponding scenario in Table 2. The regression results are then used to predict taxi-out time for the simulation runs of the same scenario in the other two weeks. Each week is

chosen as the “regression week” and the resulting estimated mode is used to predict the other two weeks. Thus, each week’s taxi-out time is predicted twice—based on the regressions estimated in the other two weeks. In total, therefore, we have six permutations of predictions weeks for each scenario. RMS errors, inequality coefficients and their proportions are calculated for each permutation and scenario, and the results averaged by scenario. The results are summarized in Table 3.

Table 3. Results on Taxi-out Time Prediction -Realization Analysis

Scenario	RMS Error (min)	Inequality Coefficient (U)	Inequality Proportions		
			U_m	U_s	U_c
High Traffic, Baseline	4.30	0.26	0.035	0.202	0.760
High Traffic, Advisory	0.96	0.09	0.020	0.048	0.928
Medium Traffic, Baseline	2.77	0.21	0	0.27	0.72
Medium Traffic, Advisory	1.07	0.10	0.173	0.062	0.758

At the high traffic level, with advisory, absolute and relative prediction errors are reduced by 77% and 65% respectively. At the medium traffic level, these reductions are 61% and 52% respectively. Imperfect co-variation is the largest source of the relative error, accounting for over 70% of the relative prediction error in all cases. Difference in variation is also a significant contributor in the baseline scenarios, but a smaller one in the SARDA scenarios. Error in central tendency is normally a very small source of prediction, with the one exception of the medium traffic, advisory scenario. It should be remembered that inequality coefficients represent fractions of the total relative prediction error. Thus even in the medium traffic, advisory scenario, the error in central tendency prediction causes a relative prediction error of less than 2%.

4.2.3 Variation in Taxi-out Time due to Controller Effect

In 4.2.1, we use observable aircraft count and taxi distance variables to explain variation in taxi-out time without explicitly considering the effect of differences among controllers. Such differences clearly influence airport surface operations. For instance, if there is a taxi conflict at an intersection, or contention between a departure and a crossing arrival for a runway, different ground controllers or local controllers may apply different criteria in deciding which aircraft gets priority. In other cases, controllers may simply exhibit different levels of skill in working with surface traffic. One would expect controller differences to be more important in the baseline runs, since the SARDA advisories sharply limit controller influence. In this section, we test this hypothesis, by modifying our taxi-out time models to include controller random effects. The experiment includes six different pairs of ground and local controllers; thus there are six different random effects in any given model.

To estimate controller random effects, we employ linear mixed modeling (Tang et al., 2012). These models include the same previous count and distance variables as the ones discussed above, but also include a controller random effect by assuming the coefficients, b_j in (8) are normally distributed with mean 0. The specification of the linear mixed model is thus:

$$Y_{k,j} = \beta_0 + \sum_{i=1}^4 \beta_{D,i} \cdot D_{i,k,j} + \sum_{i=1}^4 \beta_{N,i} \cdot N_{i,k,j} + b_j + \varepsilon_{k,j} \quad (8)$$

where, $Y_{k,j}$ is the taxi-out time for the k th observation in the j th group, with the group defined by the pair of controllers; $D_{i,k,j}$ and $N_{i,k,j}$ are the fixed-effect distance and count variables for observation k in group j ; $\beta_{D,i}$ and $\beta_{N,i}$ are the fixed-effect coefficients, which are identical for all groups; b_j is the random-effect coefficient for group j , assumed to be normally distributed with mean 0. The random effects, therefore, vary across controller pairs but are constant for each pair; $\varepsilon_{k,j}$ is the error for observation k in group j . This error is assumed to be normally distributed with mean 0, and to vary across all observations.

The coefficients for the random effect models are shown in Table 4. Compared to the baseline, there is 80-85% less variation in taxi-out time due to the controller effect in the advisory case, observed by comparing the random-effect variances with advisories and under the baseline. Thus SARDA substantially reduces the contribution of controller differences to taxi-out time variation. On the other hand, in both the baseline and advisory cases, the variances of the controller effect are very small compared to the residual variances. Moreover, the difference in residual variance between the baseline and advisory models is larger than the difference in the controller effect variance. Thus while SARDA does reduce the impacts of controller team differences, this effect is inconsequential compared to its impact on residual variance. These results also show that controller effects can be safely excluded from the taxi-out time models (Starkweather, 2010).

Table 4. Results on Taxi-out Time (second) Linear Mixed Modeling, Random Controller Effects

	High Traffic				Medium Traffic			
	Baseline		Advisory		Baseline		Advisory	
	Estimate	St. err.	Estimate	St. err.	Estimate	St. err.	Estimate	St. err.
Intercept	-4.82	314.43	42.70	76.36	93.72	238.23	0.14	94.53
# in the ramp	45.29	6.24***	5.93	2.63*	32.40	7.09***	0.22	2.94
# on the taxiway	35.69	4.16***	6.23	1.11***	19.24	3.02***	9.76	1.29***
# in the queue area	6.91	3.88	5.69	1.71**	3.99	3.08	4.76	2.07*
# in the arrival area	2.83	10.64	9.36	2.41***	18.64	8.78*	4.49	3.17
Ramp distance	0.90	0.08***	0.37	0.02***	0.39	0.06***	0.33	0.02***
Taxiway distance	0.11	0.02***	0.15	0.00***	0.13	0.02***	0.15	0.01***
Queue area distance	0.21	0.88	0.37	0.21	0.51	0.65	0.44	0.26
Runway distance	0.44	1.67	-0.46	0.40	-0.81	1.23	-0.40	0.50
Number of observation	535		542		423		418	
Random-effect variance	834.37		126.08		1660.9		334.42	
Residual variance	63635.35		3112.16		25431.1		3263.61	

4.3 Predictability in Take-off Sequence

Departure runways at busy airports often have queues and in some cases, as in DFW, specific runway queueing areas are created on the airport surface. When such queues are present, sequence predictability is another relevant predictability metric. If pilots can know their place in the departure sequence sooner, they can better prepare for take-off. One way to gauge departure sequence predictability is to assess the correlation between this sequence and some other sequence that is known earlier. Two such sequences are those for entry into the queue area, and for estimated departure sequence based on scheduled taxi-out time. The sequence predictability is higher when there is stronger

correlation between these sequences. Here, we compare the sequence matches between the baseline and advisory scenarios.

The sequence of the queue entry may not be the same as the departure sequence given multiple lanes in the queue area, as shown in Figure 1. In the baseline case, controllers use their own judgment to sequence departures in order to make efficient use of the runway, which may compromise predictability. The SARDA algorithms share similar objectives: maximizing throughput and minimizing delay (Malik et al., 2010; Jung et al., 2011; Gupta et al., 2012). There is also no attempt to maintain the match in the sequences in the queue area. It is therefore interesting to see how the sequence predictability changes with the use of SARDA advisories. Sequence predictability in each run is defined using the Spearman's coefficient (Blalock, 1979), which measures the sequence correlation between the queue entry and the aircraft take-off roll. Mathematically, the metric can be written as:

$$\alpha_{p,s} = 1 - \frac{6 \sum_{i=1}^N (R_i^{queue} - R_i^{roll})^2}{N(N^2 - 1)} \quad (9)$$

where, $\alpha_{p,s}$ is sequence predictability for each simulation run; N is number of aircraft in the run; R_i^{queue} is the rank of aircraft i at the queue entry time among all the aircraft; R_i^{roll} is the rank of aircraft i at the take-off roll time among all the aircraft. The value of the metric will be 1 if the sequences are in complete agreement.

Using this formula, predictability is calculated for the advisory and baseline runs at the medium and high traffic levels. The results are summarized in Table 5, where the simulation runs are paired by taxi-out schedules and controllers. In other words, the two sequence predictability estimates in each row are for the two runs—one baseline run and one advisory run, where taxi-out schedules are the same and the local and ground controllers are the same. As mentioned earlier, there are four identical runs: m1, m2, h3 and h4 and six pairs of controllers. Therefore, there are 12 pairs of simulation runs at each traffic level.

As shown in the table, the sequence predictability is high in both the baseline and the advisory case. On average, rank correlation is larger with advisories, indicating higher sequence predictability. The differences are highly significant, based on a paired t-test. The improvement in predictability is more obvious at high traffic level. Taking a closer look at the data, we find that one reason for a lower rank correlation in the baseline runs is the TMI flights. Without advisories, controllers tend to taxi out TMI flights earlier and have them wait in the queue well in advance to make sure the TMI flights will make the assigned time window. This strategy reduces the rank correlation between queue entry time and take-off roll time. Another reason for the better performance of the advisory case is the reduced number of aircraft in the queue area with the use of SARDA. In the experiment, it was observed that the number of aircraft in the departure queue rarely exceeded six in the advisory case, whereas in the baseline case queue sizes of up to 16 were observed. Smaller overall number of aircraft in the runway queue potentially leads to less possibility of deviation from first-in-first-out performance in the queue area, and hence better sequence predictability in the advisory case.

Table 5. Results on Sequence Predictability in the Queue Area

Runs	Medium Traffic Level			Runs	High Traffic Level		
	Baseline	Advisory	Improvement		Baseline	Advisory	Improvement
1	0.984	0.997	0.013	1	0.974	0.996	0.022
2	0.992	0.997	0.005	2	0.986	0.997	0.011
3	0.988	0.997	0.009	3	0.926	0.998	0.072
4	0.982	0.995	0.013	4	0.938	0.999	0.061
5	0.991	0.993	0.002	5	0.983	0.996	0.013
6	0.985	0.993	0.008	6	0.978	0.998	0.02
7	0.993	0.998	0.005	7	0.936	0.999	0.063
8	0.978	0.995	0.017	8	0.967	0.997	0.03
9	0.992	0.998	0.006	9	0.985	0.998	0.013
10	0.989	0.996	0.007	10	0.980	0.997	0.017
11	0.984	0.997	0.013	11	0.929	0.999	0.07
12	0.984	0.998	0.014	12	0.950	0.999	0.049
		Average	0.009			Average:	0.037
Test Statistics (t-test):			7.060	Test Statistics (t-test):			5.237
p-value (one-tailed):			<0.0001	p-value (one-tailed):			0.0001

Sequence predictability may also be assessed at a strategic level. Assuming no congestion at the airport, we could estimate the scheduled take-off roll time for aircraft, using the scheduled pushback time plus unimpeded taxi-out time. We use scheduled take-off roll time instead of scheduled pushback time because the unimpeded taxi time for aircraft from different terminals can be very different. To assess the impact of SARDA on the ability to predict take-off roll sequence from the scheduled take-off roll sequence, the sequence predictability between the two take-off roll times is estimated following the same method as before. The results are summarized in Table 6. In this case, the difference between the baseline and advisory runs is statistically insignificant given large p-values. It is notable that in the baseline operation, flight operators are motivated to push back from the gate as close to their scheduled pushback time as possible in order to maintain their place in the departure sequence; in the advised operations, collaborative gate-holding is incorporated which could potentially deviate actual take-off roll sequence from scheduled take-off roll sequence. However, the results show that there is no significant difference in the sequence predictability performance resulting from the use of SARDA.

Table 6. Results of Sequence Predictability between Scheduled and Actual Take-off Time

Runs	Medium Traffic Level			Runs	High Traffic Level		
	Baseline	Advisory	Improvement		Baseline	Advisory	Improvement
1	0.976	0.961	-0.015	1	0.965	0.951	-0.014
2	0.972	0.954	-0.018	2	0.934	0.957	0.023
3	0.962	0.975	0.013	3	0.924	0.933	0.009
4	0.957	0.971	0.014	4	0.927	0.940	0.013
5	0.984	0.956	-0.028	5	0.966	0.952	-0.014
6	0.966	0.952	-0.014	6	0.959	0.956	-0.003
7	0.989	0.978	-0.011	7	0.929	0.935	0.006
8	0.965	0.975	0.01	8	0.943	0.940	-0.003
9	0.989	0.959	-0.03	9	0.944	0.954	0.01
10	0.981	0.952	-0.029	10	0.928	0.960	0.032
11	0.967	0.973	0.006	11	0.920	0.925	0.005
12	0.936	0.947	0.011	12	0.914	0.938	0.024
		Average	-0.008			Average:	0.007
Test Statistics (t-test):			-1.511	Test Statistics (t-test):			1.755
p-value (one-tailed):			0.159	p-value (one-tailed):			0.107

5. Entropy of the Airfield State

Entropy has been used to measure uncertainty and unpredictability since its introduction into information theory by Shannon (1948). He defined the entropy of an event, x , that can take a finite set of values as:

$$H(x) = - \sum_{i=1}^n p_i \cdot \log_2(p_i) \quad (10)$$

where, p_i is the probability that the event i takes value; and $0 \cdot \log_2 0$ is set to its limiting value of 0. Since probability is between 0 and 1, the natural logarithm of the probability will never exceed 0 and thus entropy is always equal to or larger than 0.

Entropy can also be determined for homogeneous ergodic Markov chain (MC) process with finite state space (Mcmillan, 1953; Ciuperca and Girardin, 2005). The entropy rate could be measured in two ways:

$$H = - \sum_{i=1}^n \pi_i \cdot \log_2(\pi_i) \quad (11)$$

or

$$H = - \sum_{i=1}^n \pi_i \sum_{j=1}^n p_{ij} \log_2(p_{ij}) \quad (12)$$

where, π_i is the stationary probability for state i ; p_{ij} is the transition probability from state i to state j . Equation 11 measures the entropy of the system being at different states without considering transitions between states. Equation 12 measures the average entropy rate of the next move from each state, with the average weighted by the stationary probability of being at that state.

There are various ways of defining MC processes in the context of airport surface operations. For example, the progress of a flight from gate pushback to take-off can be depicted as a sequence of states demarcated by the events considered in Section 4. If the states are observed every minute, then the flight might be in the ramp area state for the first five steps after which its state transitions to the taxiway area. However, this MC process is not suitable for entropy rate analysis because the states are not recurrent and the MC process thus cannot be stationary. To address this, the MC processes considered in our study define the states by the numbers of aircraft in different areas of the airfield. In particular, we focus on numbers of aircraft on the taxiway and in the queue area. We consider these as two separate MC processes for reasons of tractability and because the two areas are controlled by different controllers. For each MC process, we used the simulation data to estimate the following transition matrix:

$$P = \begin{bmatrix} p_{1,1} & \cdots & p_{1,n} \\ \vdots & \ddots & \vdots \\ p_{n,1} & \cdots & p_{n,n} \end{bmatrix} = \begin{bmatrix} \frac{N(1,1)}{\sum_{i=1}^n N(1,i)} & \cdots & \frac{N(1,n)}{\sum_{i=1}^n N(1,i)} \\ \vdots & \ddots & \vdots \\ \frac{N(n,1)}{\sum_{i=1}^n N(n,i)} & \cdots & \frac{N(n,n)}{\sum_{i=1}^n N(n,i)} \end{bmatrix} \quad (13a)$$

$$p_{i,j} = \Pr(S_{t+1} = j | S_t = i) \quad (13b)$$

where, P is the transition matrix; $p_{i,j}$ is the transition probability from state i (i aircraft) to state j (j aircraft); $N_{i,j}$ is the number of transitions from state i to state j ; S_t is the state at time step t , defined by the number of aircraft in the area—either the taxiway or the queue area at the time step t . The stationary distribution Π is calculated using the following equations:

$$\Pi = [\pi_1, \dots, \pi_n] = \Pi \cdot P \quad (14a)$$

$$\sum_{i=1}^n \pi_i = 1 \quad (14b)$$

Where, π_i is the stationary probability for state i , and the sum of the stationary probabilities for all the states is equal to 1. It takes time at the beginning of each simulation run for the system to reach its steady state. At the end of the simulation, traffic decreases because no aircraft is pushing back. An eligible MC process for entropy analysis must be stationary (Ciuperca and Girardin, 2005). For this reason, we drop the observations from the transition periods and only keep the observations when the systems are judged to be in steady state for entropy analysis. The time step is set as one minute for the MC processes. Considering that only one aircraft can be released to the runway for any given time and the required separation between take-offs, the count of aircraft in the queue area can drop by no more than two between successive time steps. Non-zero transition probabilities are therefore distributed close to the diagonal entries in the transition matrices for the queue area MC process. On the taxiway, there are two types of aircraft: departure aircraft released from their spot and arriving aircraft crossing to go to their gates. The count of aircraft on the taxiway would go down when a departure aircraft enters the queue area or an arriving aircraft enters the ramp area. Because there are multiple possible sources and sinks, there should be more possible state transitions for aircraft counts on the taxiway as compared to queue area counts. The transition matrices for the taxiway MC process therefore have more non-zero entries.

The entropy analysis is performed for each combination of traffic level and advisory situation. The entropy values are summarized in Table 7. Tables 8 and 9 present the stationary probabilities of number of aircraft in the queue area and on the taxiway.

Table 7. Entropy Rates of the Markov Chain Processes in the Queue Area and on the Taxiway

	Without Transition (Equation 11)		With Transition (Equation 12)	
	Queue Area	Taxiway	Queue Area	Taxiway
High Traffic, Baseline	3.430	3.368	2.079	2.633
High Traffic, Advisory	2.210	2.918	1.830	2.615
Medium Traffic, Baseline	3.292	3.126	2.042	2.420
Medium Traffic, Advisory	2.203	2.799	1.853	2.433

The first two columns of entropy values are calculated for the aircraft count distributions in the queue area and on the taxiway based on Equation 11. Use of SARDA advisories brings down the entropy values in both areas at both traffic levels. The fewer

the possible states and the more concentrated the distribution of the states, the lower the entropy values. As shown in Table 8, with SARDA, the number of aircraft in the queue area remains under seven and the probability distribution of the count is concentrated around four. In the baseline, the number of aircraft in the queue area goes to as high as 16 at high traffic level and 14 at medium traffic level. Furthermore, the distribution of the number of aircraft is more dispersed in the baseline scenarios. As shown in Table 9, use of SARDA reduces the maximum number of aircraft on the taxiways by two and three at the high and medium traffic levels respectively. The impact of the advisory on the concentration of the distribution is not significant. As a result, the decrease in taxiway entropy that results from using SARDA is less than the decrease in queue area entropy. This implies that SARDA is probably more effective in maintaining a steady and predictable workload for the local controller, who is responsible for the queue area traffic. In the next section, this conjecture is verified by the survey results.

The last two columns of entropy rates in Table 7 are calculated considering transitions between states based on Equation 12. More possible states do not necessarily lead to higher entropy when it is calculated using this equation. For instance, assume that the transition matrix is such that each row contains the same set of non-zero probability values, but arranged differently across the columns. Then the entropy rate would be the same regardless of the number of possible states and their stationary probabilities. As seen in the table, with the advisory entropy rates are slightly lower for the queue area but virtually identical for the taxiways. This indicates that advisory does not greatly improve the predictability of the minute-to-minute evolution of the airfield.

Table 8. Stationary Probabilities of Number of Aircraft in the Queue Area

Number of aircraft	High Traffic, Baseline	High Traffic, Advisory	Medium Traffic, Baseline	Medium Traffic, Advisory
1	0	0.015	0.019	0.048
2	0	0.078	0.042	0.062
3	0	0.191	0.052	0.292
4	0.037	0.346	0.093	0.308
5	0.066	0.294	0.121	0.243
6	0.062	0.067	0.134	0.047
7	0.116	0.010	0.189	
8	0.111		0.164	
9	0.118		0.071	
10	0.156		0.065	
11	0.131		0.034	
12	0.091		0.011	
13	0.047		0.003	
14	0.036		0.003	
15	0.021			
16	0.008			

Table 9. Stationary Probabilities of Number of Aircraft on the Taxiway

Number of aircraft	High Traffic, Baseline	High Traffic, Advisory	Medium Traffic, Baseline	Medium Traffic, Advisory
1	0.017	0.003	0.038	0.031
2	0.088	0.028	0.127	0.065
3	0.117	0.127	0.183	0.091
4	0.129	0.176	0.188	0.200
5	0.152	0.168	0.128	0.216
6	0.092	0.179	0.103	0.210
7	0.107	0.157	0.071	0.128
8	0.088	0.096	0.102	0.049
9	0.107	0.055	0.029	0.010
10	0.058	0.008	0.019	
11	0.031	0.003	0.006	
12	0.007		0.006	
13	0.005			

6. Controller Survey

As mentioned, six recently retired controllers participated in the simulation experiments over a period of three weeks, with each pair running scenarios for an entire week. There were a total of 48 data-collection runs in the three weeks. As part of the simulation experiment, participating controllers were asked to complete surveys on their experience using SARDA. Two types of surveys were presented to the controllers: post-run surveys, which were given to the controllers at the end of each run; and post-participation surveys, which were given to the controllers at the end of each simulation week. The survey questions of both are presented in the Appendix. Both surveys included questions pertaining to predictability.

6.1 Post-run Survey

The post-run survey included three questions related to predictability, as shown in Table 10. The first question concerned controllers' perceptions of workload stability in the run; when there are less unexpected surges, the workload is more stable and hence predictable. The other two questions pertained to accommodating TMI flights. In the simulation, controllers were asked to take off TMI flights within a one-minute window of the TMI time. Therefore, the TMI flight take-off times requires more predictability.

The survey results are separated into two sets: responses for baseline runs and responses for advisory runs. The Wilcoxon Matched-pair Signed-ranks Test (Blalock, 1979) is selected to assess whether or not there is a significant difference between the responses in the two sets, where the responses are paired with pushback schedules and controllers. The null hypothesis assumes there were no differences between the mean responses of the two sets. The alternative hypothesis assumes SARDA improved the average performance. The test results are summarized in the following table.

Table 10. Post-run Survey Results, with Significance Level as 0.05

Questions	Ground controller	Local controller
1. Unexpected surges in workload	No difference	Significant reduction
2. Frequency in delaying non-TMI flights to make sure that TMI flights could make their windows	No difference	No difference
3. Extra attention required for managing TMI flights	Significant reduction	Significant reduction

The results indicate that under advisory there are fewer unexpected surges in workload for local controllers but not for ground controllers. SARDA does not have a significant impact on the need to delay non-TMI flights in order to accommodate TMI ones. However, these advisories reduce the attention required for managing TMI flights for both positions.

6.2 Results of Post-participation Survey

In the post-participation questionnaire, controllers were asked to assess predictability-related impacts of SARDA on handling hypothetical off-nominal events, avoiding queue spillovers and providing accurate information to pilots. The off-nominal events involve aspects of airport surface operations that were not incorporated into the simulation, but which can be disruptive and thus undermine predictability. The complete set of the off-nominal events can be found in the appendix. Examples include a departing aircraft that needs to return to the gate, or a change in runway configuration. Two features of the impacts were surveyed. First, controllers were asked to assess the degree of impact of SARDA on their ability to handle the event. Then they were asked to assess the importance of the potential improvement in handling that event.

The survey results are summarized and shown in a 2-dimensional plot in Figure 3, where both the average degree and the average importance of each impact are assessed. The degree of the impact is assessed using a scale of 1 to 10, where 1 means much worse with SARDA and 10 means much better with SARDA. An average grade above 5.5 indicates that performance is improved with SARDA. Figure 3 shows that in the view of controllers, SARDA promises some improvement in each of the aspect assessed. Further, controllers believe that the greatest and most impactful improvements from SARDA would occur when there is a change in runway configuration or departure routes. Compared to these, reducing confusion about call signs and guide in runway crossing are also stated to be important, but SARDA may not offer substantial improvement in these aspects.

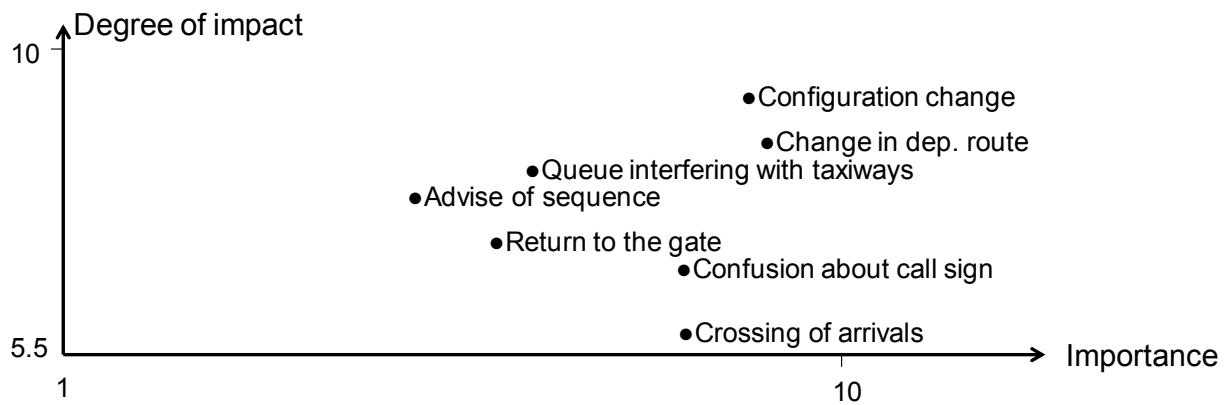


Figure 3. Post-participation Survey Results

7. Conclusions

Past evaluations of airport surface operations automation technologies have focused on capacity utilization, delay mitigation and fuel efficiency impacts. Predictability, while recognized as an important operational performance goal, has received little attention. The lack of predictability metrics hampers our ability to assess the predictability impact of automation tools. This research fills in the gap of performance measurement by defining and quantifying predictability in the context of airport surface operation management. Using results from a SARDA human-in-the-loop simulation conducted at NASA Ames' Future Flight Central, we present a comprehensive assessment of the predictability impacts of airport surface automation.

Unpredictability in total taxi-out time—from scheduled pushback to actual take-off—is measured as the integration of standard deviation of remaining taxi-out time over time. The results show that SARDA reduces unpredictability by 46% and 39% under high and medium traffic levels, respectively. The reductions happen because variability in total taxi-out time is somewhat smaller, and variability in remaining taxi-out time declines much faster, when SARDA advisories are followed.

Regression models are constructed to explain and predict the variability in actual taxi-out time—from actual pushback to actual take-off. To make the model predictive, we only use variables with values known at the actual pushback time: nominal taxi distances and numbers of aircraft in different airfield areas. When SARDA is used, absolute prediction errors and relative prediction errors are reduced by 77%/61% and 65%/52% respectively at high/medium traffic levels. Linear mixed models are constructed to study the contributions of differences between controllers to taxi-out time variation. While SARDA substantially reduces the impact of controller differences, this effect is inconsequential compared to its impact on residual variance.

Sequence predictability is measured by comparing the actual take-off sequence to the queue entry sequence and the scheduled take-off sequence. The improvement in sequence predictability in the queue area is modest but highly significant with advisories, whereas the sequence predictability based on scheduled take-off sequence is about the same with and without SARDA advisories.

Uncertainty in the number of aircraft on the taxiway and in the queue area is measured using entropy rates of Markov Chain processes, where the states are defined as the numbers of aircraft in these areas at each time step. For the queue area, SARDA reduces the number of possible states and makes the stationary distribution of states more concentrated. These two impacts bring down the entropy values of the possible states. For the taxiway area, the impact of SARDA on the entropy rate of the distribution of states is not substantial. For both areas, the entropy rates are similar with and without advisories when transitions between states are considered. This indicates that SARDA does not improve the predictability of how many aircraft there would be in the system in the next time step based on the number of aircraft in the current step.

Surveys are conducted to investigate controllers' perceptions regarding predictability-related impact. The surveys reveal that SARDA reduces the attention required for Traffic Management Initiative flights for both ground and local controllers, and unexpected surges in workload for the local controller. Controllers also expect that SARDA advisories would yield major benefits in managing traffic during runway configuration

changes and departure route changes, although these were not part of the simulation because the current version of SARDA does not include these capabilities.

Summing up these results, we find that SARDA improves predictability in airport surface operations based on a wide range of metrics and criteria. Two major and closely related reasons for these impacts are the pushback advisories, which enable tactical gate-holding, and the reduction of aircraft in the queue area. The former greatly improves foreknowledge of take-off time, while the latter lessens the variability in queue area occupancy. The contributions of other SARDA advisories, such as those for departure sequences and runway crossings, are less clear.

In assessing the predictability impacts of SARDA, this research also elucidates what predictability means in the context of airport surface operations. When the necessary data are available, the predictability metrics proposed in this paper could be applied to assess the predictability impact of other automation tools, at different levels of automation, at other airports, and in field deployments as well as simulations. Through such efforts, claims of predictability improvements provided by automation technologies can become clear statements subject to rigorous analysis, rather than vague and unverifiable sentiments. We thus provide a new perspective on evaluating the impacts of emerging technologies on airport surface operations performance.

While our metrics are widely applicable, they do not provide a complete picture of predictability performance in situations where intent and control information are available to system users. In these cases, information about specific taxi-out times, departure sequences, and so on, could be provided to individual flights even though dispersion and imperfect correlation are still present. In such cases, our metrics might be supplemented by others that more precisely reflect what agents know and when they know it. This is an area for future research.

An obvious follow-up question pertains to the value of these predictability improvements. Monetizing the predictability improvements is much less straightforward than monetizing, say, fuel savings. Certain predictability-related benefit mechanisms, such as allowing more single-engine taxiing and improved adherence to 4-D trajectories, were cited earlier. There may be other benefits of a psychological or human performance nature, such as smoother controllers' workload. Additional research is needed to assess the value of all of these benefits in monetary terms. In the meantime, it should be remembered that the community has already recognized that predictability is an important performance goal.

Acknowledgments

The study is supported by NASA Aligned Research Program Award. We would like to thank Frank Ketcham, pilot with Delta Airlines, for his valuable comments on aircraft surface operations. We would also like to thank Victoria L. Dulchinos, Miwa Hayashi and Ty Hoang at NASA Ames Research Center for their support in the research process. Last but not least, special thanks go to Professor Rhonda Righter at University of California, Berkeley, for her comments on Markov Chain Process.

References

- Air Traffic Services Performance Focus Group (ATSPFG), CNS/ATM Focused Team, 1999. Report: Airline Metric Concepts for Evaluating Air Traffic Service Performance.
- Ball, M.O., Hoffman, R.L., Knorr, D., Wetherly, J., Wambsganss, M., 2000. Assessing the benefits of collaborative decision making in air traffic management. 3rd USA/Europe Air Traffic Management Research and Development Seminar, Naples, Italy, 13 – 16 June 2000.
- Bates, J., Polak, J., Jones, P., Cook, A., 2001. The valuation of reliability for personal travel. *Transportation Research Part E*, 37, 191 – 229.
- Blalock, H.M. 1979. Social Statistics, 2nd Edition. McGraw-Hill, Inc. USA.
- Borokhov, P., Blandin, S., Samaranayake, S., Goldschmidt, O., Bayen, A., 2011. An adaptive routing system for location-aware mobile devices on the road network. 14th IEEE Conference on Intelligent Transportation Systems, Washington, D.C., 5 – 7 October 2011.
- Bradford, S., Knorr, D., Liang, D., 2000. Performance measures for future architecture. 3rd USA/Europe Air Traffic Management Research and Development Seminar, Naples, Italy, 13 – 16 June 2000.
- Brinton, C., Provan C., Lent, S., Prevost, T., Passmore, S., 2011. Collaborative departure queue management. 9th USA/Europe Air Traffic Management Research and Development Seminar, Berlin, Germany, 14 – 17 June 2011.
- Chauhan, A., 2010. Modeling and Predicting Taxi Time at Airports. Master Thesis, University of South Florida.
- Ciuperca, G., Girardin, V., 2005. On the Estimation of the Entropy Rate of Finite Markov Chains, Proceedings of the International Symposium on Applied Stochastic Models and Data Analysis.
- Clewlow, R.R., Simaiakis, I., Balakrishnan, H., 2010. Impact of arrivals on departure taxi operations at airports. 10th American Institute of Aeronautics and Astronautics Aviation Technology, Integration, and Operations Conference, Fort Worth, Texas, 1 – 15 September 2010.
- Cohen, H., Southworth, F., 1999. On the measurement and valuation of travel time variability due to incidents on freeways. *Journal of Transportation Statistics* 2 (2), 123 – 131.
- Cover, T.M., Thomas, J.A., 1991. Element of Information Theory, John Wiley & Sons, Inc. USA.
- Ettema, D., Timmermans, H., 2006. Costs of travel time uncertainty and benefits of travel time information: conceptual model and numerical examples. *Transportation Research Part C* 14, 335 – 350.
- Eurocontrol, Performance Review Commission, 2012. An Assessment of Air Traffic Management in Europe during the Calendar Year 2011, Performance Review Report.
- Federal Aviation Administration (FAA), 2011. Report: Destination 2025.
- Federal Aviation Administration, 2013. http://www.faa.gov/nextgen/why_nextgen_matters/.
- Gulding, J., Knorr, D., Rose, M., Bonn, J., Enaud, P., Hegendoerfer, H., 2009. US/Europe comparison of ATM-related operational performance. 8th USA/Europe Air Traffic Management Research and Development Seminar, Napa, California, 29 June – 2 July 2009.
- Gupta, G., Malik, W., Jung, Y., 2012. An integrated collaborative decision making and tactical advisories concept for airport surface operations management. 12th American Institute of Aeronautics and Astronautics Aviation Technology, Integration, and Operations Conference, Indianapolis, Indiana, 17 – 19 September 2012.
- Gupta, G., Malik, W., Tobias, L., Jung, Y.C., Hoang, T., Hayashi, M., 2013. Performance evaluation of individual aircraft based advisory concept for surface management. 10th USA/Europe Air Traffic Management Research and Development Seminar, Chicago, Illinois, 10 – 13 June 2013.
- Hao, L., Hansen, M., 2013. Flight time predictability: concepts, metrics and impact on schedule block time. Transportation Research Board, Washington, D.C., 13 – 17 January 2013.
- Hayashi, M., Hoang, T., Jung, Y., Gupta, G., Malik, W., Dulchinos, V., 2013. Usability evaluation of the Spot and Runway Departure Advisor concept in a Dallas/Fort Worth airport tower simulation. 10th USA/Europe Air Traffic Management Research and Development Seminar, Chicago, Illinois, 10 – 13 June 2013.
- Hoang, T., Jung, Y.C., Holbrook, J.B., Malik, W., 2011. Tower controllers' assessment of the Spot and Runway Departure Advisor (SARDA) Concept. 9th USA/Europe Air Traffic Management Research and Development Seminar, Berlin, Germany, 14 – 17 June 2011.
- International Civil Aviation Organization (ICAO), 2005. Report: Global Air Traffic Management Operational Concept, 1st Edition.

- Idris, H., Clarke, J.P., Bhuva R., Kang L., 2002. Queuing model for taxi-out time estimation. *Air Traffic Control Quarterly* 10(1), 1 – 22.
- Jacquet, P., Seroussi, G., Szqankowski, W., 2008. On the entropy of a hidden Markov process. *Theoretical Computer Science* 395, 203 – 219.
- Jung, Y.C., Hoang T., Montoya J., Gupta G., Malik W., Tobias, L., 2010. A concept and implementation of optimized operations of airport surface traffic. 10th American Institute of Aeronautics and Astronautics Aviation Technology, Integration, and Operations Conference, Fort Worth, Texas, 1 – 15 September 2010.
- Kho S.Y., Park, J.S., Kim, Y.H., Kim, E.H., 2005. A development of punctuality index for bus operation. *Journal of the Eastern Asia Society for Transportation Studies* 6, 492 – 504.
- Kim, A., Hansen, M., 2010. Validation of runway capacity models. *Transportation Research Record: Journal of the Transportation Research Board*, 2177, 69 – 77.
- Knorr, D., Post, J., Biros, J., 2000. The operational assessment of free flight phase 1 ATM capabilities. 3rd USA/Europe Air Traffic Management Research and Development Seminar, Naples, Italy, 13 – 16 June 2000.
- Kristof, T., Lowry, M., Rutherford, G.S., 2005. Assessing the Benefits of Traveler and Transportation Information Systems. *Research Report*.
- Kwon, J., Coifman, B., Bickel, P., 2000. Day-to-day travel-time trends and travel-time prediction from loop-detector data. *Transportation Research Record: Journal of the Transportation Research Board*, 1717, 120 – 129.
- Legge, J., Levy, B., 2008. Departure taxi time predictions using ASDE-X surveillance data. 26th International Congress of the Aeronautical Sciences, Alaska, 14 – 19 September 2008.
- Levinson, D., 2003. The value of advanced traveler information systems for route choice. *Transportation Research Part C*, 11, 75 – 87.
- Li, R., Rose, G., 2011. Incorporating uncertainty into short-term travel time predictions. *Transportation Research Part C*, 19, 1006 – 1018.
- Liu, Y., Hansen, M., In Press. Ground Delay Program Performance Evaluation. Accepted for publication in *Transportation Research Record: Journal of the Transportation Research Board*.
- Lomax, T., Schrank, D., Turner, S., Margiotta, R., 2003. Selecting Travel Reliability Measures. Texas Transportation Institute, Cambridge Systematics Inc.
- Malik, W., Gupta, G., Jung, Y.C., 2010. Managing departure aircraft release for efficiency airport surface operations. American Institute of Aeronautics and Astronautics Guidance, Navigation, and Control Conference, Toronto, Canada, 2 – 5 August 2010.
- McMillan, B., 1953. The basic theorems of information theory. *Annals of Mathematical Statistics* 24 (2), 196 – 219.
- Medina, M., Sherry, L., Feary, M., 2010. Automation for task analysis of next generation air traffic management systems. *Transportation Research Part C*, 18, 921 – 929.
- Nakahara, A., Reynolds, T.G., White, T., Maccarone, C., Dunsky, R., 2011. Analysis of a surface congestion management technique at New York JFK airport. 11th American Institute of Aeronautics and Astronautics Aviation Technology, Integration, and Operations Conference, Virginia Beach, VA, 20 – 22 September 2011.
- Nakahara, A., Reynolds, T.G., 2012. An approach for estimating current and future benefits of airport surface congestion management techniques. 12th American Institute of Aeronautics and Astronautics Aviation Technology, Integration, and Operations Conference, Indianapolis, Indiana, 17 – 19 September 2012.
- Rice, J., Zwet, E.V., 2004. A simple and effective method for predicting travel times on freeways. *IEEE Transactions on Intelligent Transportation Systems* 5(3), 200 – 207.
- Riikka, S., Paavilainen, J., 2010. Measuring railway traffic punctuality from the passenger's perspective. 12th World Conference in Transportation Research, Lisbon, Portugal, 11 – 15 July 2010.
- Shannon, C.E., 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27, 379-423, 623-656.
- Simaiakis, I., Khadilkar H., Balakrishnan, H., Reynolds, T.G., Hansman, R.J., Reilly, B., Urlass, S., 2011. Demonstration of reduced airport congestion through pushback rate control. 9th USA/Europe Air Traffic Management Research and Development Seminar, Berlin, Germany, 14 – 17 June 2011.
- Simaiakis, I., 2013. Analysis, Modeling and Control of the Airport Departure Process. Doctoral Dissertation. Massachusetts Institute of Technology.

- Starkweather, J., 2010. Linear Mixed Effects Modeling using R.
- Tang, L., Thakuriah, P., 2012. Ridership effects of real-time bus information system: a case study in the city of Chicago. *Transportation Research Part C* 22, 146 – 161.
- Taylor, M.A.P., 1982. Travel time variability- the case of two public modes. *Transportation Science* 16 (4), 507 – 521.
- Theil, H., 1966. *Applied Economic Forecasting*, North-Holland Publishing Company, Amsterdam, Netherlands.
- Zhang, X., Rice, J.A., 2003. Short-term travel time prediction. *Transportation Research Part C* 11, 187 – 210.
- Zhang, Y., Chauhan, A., Chen, X., 2010. Modeling and predicting taxi out time at airports. 4th International Conference on Research in Air Transportation, Budapest, Hungary, 1 – 4 June 2010.
- Zhang, Y., Wang, Q., 2011. Methods for determining airport unimpeded taxi times. *Transportation Research Board*, Washington, D.C., 23 – 27 January 2011.

Appendix

A. Post-run Survey Questions:

- Q1. Please rate how much you agree or disagree with the following statement: There were unexpected surges in workload. Please answer on a scale of 1 to 10 (1= Extremely disagree and 10 = Extremely agree)
- Q2. How often did you delay non-TMI flights in order to make sure that TMI flights could depart within their windows? Please answer on a scale of 1 to 10 (1= Never and 10 = Always)
- Q3. How much attention was required for managing TMI flights as compared to non-TMI flights? Please answer on a scale of 1 to 10 (1 = much less attention and 10 = much more attention)

B. Post-participation Survey Questions:

Under SARDA, there will normally be fewer aircraft on the taxiways and the queue area. Please answer the following questions:

- Q1. Will having fewer aircraft on the taxiways and queue area make it easier to handle aircraft in these locations that must return to the gate (due to a medical emergency or mechanical problem)? Please use a scale of 1 to 10, (1 = Much harder, and 10 = Much easier)
 If it is easier to handle aircraft in these locations that must return to the gate, is this an important advantage? Please use a scale of 1 to 10, (1 = Not at all important, and 10 = very important)
- Q2. Will having fewer aircraft on the taxiways and queue area lead to less confusion about the call sign of each aircraft? Please use a scale of 1 to 10, (1 = Much more confusion, and 10 = Much less confusion)
 If there is less confusion about the call sign of each aircraft, is this an important advantage? Please use a scale of 1 to 10, (1 = Not at all important, and 10 = very important)
- Q3. Will having fewer aircraft on the taxiways and queue area make it easier to advise an aircraft of its sequence? Please use a scale of 1 to 10, (1 = Much harder, and 10 = Much easier)
 If it is easier to advise an aircraft of its sequence, is this an important advantage? Please use a scale of 1 to 10, (1 = Not at all important, and 10 = very important)
- Q4. Will having fewer aircraft on the taxiways and queue area make it easier to handle airport configuration changes? Please use a scale of 1 to 10, (1 = Much harder, and 10 = Much easier)
 If it is easier to handle airport configuration changes, is this an important advantage? Please use a scale of 1 to 10, (1 = Not at all important, and 10 = very important)
- Q5. Will having fewer aircraft on the taxiways and queue area make it easier to adapt to changes in departure routes necessitated by convective weather? Please use a scale of 1 to 10, (1 = Much harder, and 10 = Much easier)
 If it is easier to adapt to changes in departure routes necessitated by convective weather, is this an important advantage? Please use a scale of 1 to 10, (1 = Not at all important, and 10 = very important)
- Q6. Will having fewer aircraft on the taxiways and queue area lead to less risk of the queue interfering with taxiways? Please use a scale of 1 to 10, (1 = Much more risk, and 10 = Much less risk).
 If there is less risk of the queue interfering with taxiways, is this an important advantage? Please use a scale of 1 to 10, (1 = Not at all important, and 10 = very important).

- Q7. Will having fewer aircraft on the taxiways and queue area make it easier to manage runway crossings of arriving aircraft? Please use a scale of 1 to 10, (1 = Much harder, and 10 = Much easier)
If it is easier to manage runway crossings of arriving aircraft, is this an important advantage? Please use a scale of 1 to 10, (1 = Not at all important, and 10 = very important)