Predictability in Airport Surface Operation Management

Yi Liu[1], Mark Hansen[2] and Ziyi Wang[3]
University of California, Berkeley, CA, 94720

Gautam Gupta[4] and Waqar Malik[5]
University of California, Santa Cruz, NASA Ames Research Center, Moffett Field, CA 94035

The performance of airport surface operations has usually been assessed with respect to delay, capacity and efficiency. Although predictability as a performance measure is recognized by stakeholders as an important goal, predictability metrics have not been defined for airport surface operations. This paper aims to fill that gap by using data from NASA's Spot and Runway Departure Advisor (SARDA) human-in-the-loop simulations in 2012 to study airport operations predictability. Using the simulation data, we measure and compare predictability on the airfield with and without SARDA from three perspectives: controllers' perspective, flight operator's perspective and traffic management perspective. The controller survey results indicate the perception that SARDA reduces controller's workload surges and has the potential to better handle off-nominal situations. By studying taxi-out time in both baseline and advisory runs, it is found that SARDA reduces variability in total taxi-out time and eliminates uncertainty in taxi-out time sooner into the taxi-out process. Moreover, SARDA enables more accurate predictions of wheels-off time through use of a linear regression model. There is no evidence indicating that SARDA causes more deviation from First-Scheduled-First-Served as compared to the non-SARDA case. Instead, SARDA improves First-In-First-Out performance in the queue area.

I. Introduction

Day-of-operation predictability allows a multitude of benefits in airport surface operations. As an example, 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 takeoff. 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. In a trajectory-based air traffic management system, investigation of predictability on the surface is not just beneficial to the airport operations but could also improve performance in the downstream. However, in the context of airport surface operation management, performance assessment has mainly been conducted with respect to delay [1-4], capacity[5-8] and efficiency[3, 9, 10]. Little work has been done to measure predictability, even though various stakeholders recognize the importance of predictability [11, 12].

National Aeronautics and Space Administration (NASA) Ames Research Center (ARC) has developed a decision support tool—Spot And Runway Departure Advisor (SARDA)—to help air traffic controllers and airline operators optimize airport surface operations. Research has shown that SARDA improves airport operational performance in terms of throughput and efficiency[3, 13]. However, the absence of predictability metrics makes it difficult to measure the potential benefit to predictability.

In this paper, we propose to bridge the gap in the literature by developing methods for measuring predictability. Using NASA's airport surface optimization project as a test bed, we investigate how airport surface traffic management strategies can be used to deliver different levels of predictability through econometric analysis of operational and behavioral data. By adding predictability into performance consideration, the research would aid

---
[1] Research Assistant, UCB, 107 McLaughlin hall, Berkeley, CA 94720. liuyisha@berkeley.edu.
[2] Professor, UCB, 114 McLaughlin Hall, Berkeley, CA 94720. mhansen@berkeley.edu.
[3] Research Assistant, UCB, 107 McLaughlin hall, Berkeley, CA 94720. wzy11@berkeley.edu.
[4] Research Scientist, UARC, Building 210, MS 210-8, Moffett Field, CA 94035. ggupta@ucsc.edu.
[5] Research Scientist, UARC, Building 210, MS 210-8, Moffett Field, CA 94035. wmalik@ucsc.edu.

traffic management decision making in a more comprehensive way. In the long term, this capability would allow us to improve the design of air traffic management decision support tools and technologies.

The remainder of the paper is as organized as follows. In section II, the human-in-the-loop simulation, which serves as the source of data, is described. More details about the simulation can be found in Ref. 13. In the same section, the controller's perception of predictability within the context of the study is discussed. In section III, taxi-out time predictability from flight operator's perspective is discussed. In section IV, from a traffic management perspective, predictability performance when SARDA is used to provide decision support is compared to that when SARDA is off. Finally, the paper concludes in section V.

II. Controller Perceived Predictability

NASA developed SARDA to provide controllers with timing advisories from the ramp area, along the taxiway onto the departure runway for departure aircraft[3, 14, 15, 16]. The objectives of SARDA are to alleviate traffic congestion on the airport surface and reduce fuel burn. In May 2012, NASA conducted a human-in-the-loop simulation evaluation of SARDA over 3 weeks[13]. The experiment was conducted for surface operations on the east side of Dallas-Fort Worth Airport (DFW) with two levels of traffic density: medium level with 40 departure aircraft in 50 minutes and high level with 50 departures in 50 minutes. To compare the difference in operational performance with and without SARDA, the team ran two sets of simulation runs each week: 8 advisory runs with SARDA on and 8 baseline runs with SARDA off, with half of the runs for each traffic density. In the advisory runs, controllers were asked to follow SARDA's advisories.

A total of 6 retired DFW tower controllers participated in the simulation. Surveys were designed to assess controller's perception of improvement due to SARDA, with many improvements being indirect measures of predictability. For instance, unexpected surges in workload can be linked to decreased predictability. Two types of surveys were presented to controllers: post-run surveys, which were given to the controllers at the end of each run; post-study surveys, which were given to the controllers at the end of each simulation week. The survey questions of both are presented in the Appendix.

A. Post-run Questionnaire

The 3 post-run survey questions were designed to assess controller perception of workload stability in the run; when there are less unexpected surges, the workload is more stable and hence predictable. The survey questions also assessed the extra attention required for managing flight with Traffic Management Initiative (TMI) restrictions compared to non-TMI flights. In the simulation, controllers were asked to take off TMI flights within a ± 1-minute window of the TMI time. Therefore, the TMI flight take-off times should be more predictable compared to the schedule take-off times. However, this increased predictability could come at the expense of extra attention from the controller, and we investigate how much more attention is paid by controllers to control TMI flights compared to non-TMI flights. The survey results are separated into two sets: responses for baseline runs and responses for advisory runs. Wilcoxon Matched-pair Signed-ranks Test is selected to assess whether or not there is significant difference between the mean responses in the two sets, where the responses are paired with traffic scenario and controller. The null hypothesis assumes there were no differences between the mean responses of the two sets. The test results are summarized in Table 1.

Table 1. Post-run survey results, with significance level as 0.05

Questions	Ground controller	Local controller
1. Unexpected surges in workload	No difference	significant difference
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 difference	Significant difference

The results indicate that under advisory there were fewer unexpected surges in workload for Local Controllers (responsible for the runways including take-offs, landings and crossings) but not for Ground Controllers (responsible for taxiway movements). Advisory did not affect prioritizing the taxi-out sequence of aircraft at the significance level. However, SARDA reduced attention required for managing TMI flights at both positions.

B. Post-study Questionnaire

In the post-study 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. Two features of the impacts were surveyed. First, controllers were asked to assess the degree of impact of SARDA

2
American Institute of Aeronautics and Astronautics

on the event. Then they were asked to assess the importance of the potential improvement in handling that event. The impact of SARDA could be good or bad, but the importance of the impact is assessed assuming improvement.

The survey results are summarized and shown in a 2-dimensional plot in Fig. 1, 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. If the average grade is above 5.5, then it means the performance is improved with SARDA. With all the points located above the horizontal axis, it indicates that SARDA has potential in improving the predictability-related performance. Further, controllers would appreciate the taxi advisories the most when there is change in runway configuration or departure routes. These two potential improvements are also considered the most important by controllers. Compared to these, reducing confusion about call signs and guide in runway crossing are also stated to be important, but SARDA does not improve these performances in a substantial way.
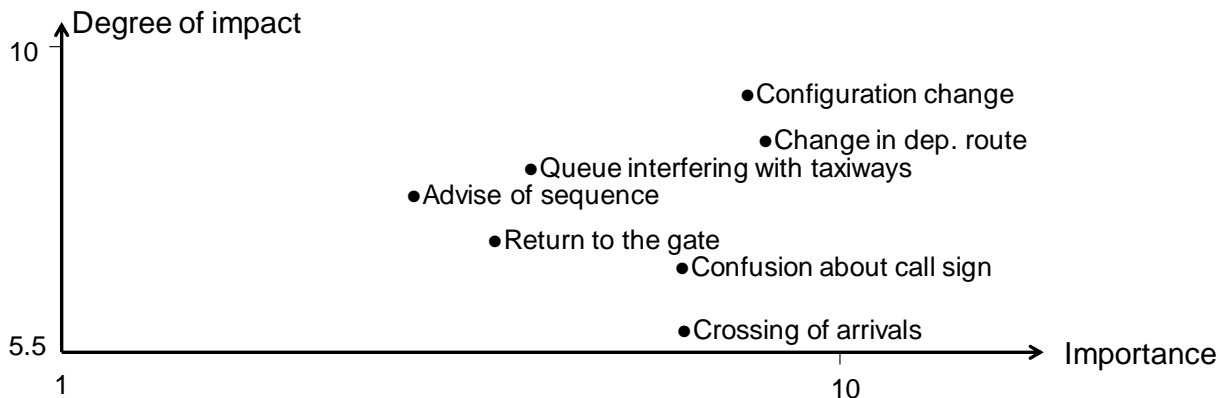

Figure 1. Post-study survey results

III.   Predictability from Flight Operator's Perspective

Simulation data from 48 runs are categorized into 4 cases based on the traffic density and advisory condition: Medium traffic with baseline, medium traffic with advisory, high traffic with baseline and high traffic with advisory. Each case was run 6 times with SARDA and 6 times in baseline (no SARDA). Based on these data, predictability from the flight operator perspective is studied in two ways. In Section A, we compare the predictability performance with and without advisories by investigating unpredictability in total taxi-out time, from scheduled pushback time to actual wheels-off time. In Section B, we analyze predictability in taxi-out time after pushback, i.e. from actual pushback to actual wheels-off time, using linear regression analysis.

A. Unpredictability in Total Taxi-out Time

1. Unpredictability Metric

Higher predictability in taxi-out time is equivalent to better prediction in the wheels-off time, which will benefit the operations both on the airfield and in the downstream of the network. In the vast majority of existing literature, predictability is defined by measuring the variation in the system experienced by the users[17, 18]. Less variability in taxi-out time means higher predictability in operations. However, accuracy in information alone might not provide a full picture of predictability.

Assume flights A and B are taxiing out at an airport. Flight A is pushing back from the gate and its wheels-off time is estimated with 1-minute standard deviation. Flight B is entering the queue area and its wheels-off time is also estimated with 1-minute standard deviation. Following the definition in the literature, the level of predictability in taxi-out time for the two flights would be considered similar. However, the error information is more helpful to flight A since the information is provided at an earlier stage and pilots have more time to adjust operations to the error in estimation. On the contrary, pilots of Flight B might have started the second engine to prepare for departure before entering the queue area. The information provided at the queue entry time is then less useful for pilots of flight B because of the late notification. Cases like this motivate the consideration of timeliness of information in the predictability evaluation.

When taxiing out, pilots want to know the wheels-off time in advance so that they can start the second engine optimally and get ready for takeoff. Generally, uncertainty in wheels-off time estimation decreases with time in the taxi-out process. Timely information permits more time of operation adaptation but could be less accurate. Late

3
American Institute of Aeronautics and Astronautics

notification is more accurate but leaves operators less ability to plan the operations. Therefore, a comprehensive discussion about predictability must consider both notions: accuracy and timeliness. In other words, information is most unpredictable if it is inaccurate and not timely. Following this, unpredictability is formulated as:

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

where, t_s is the scheduled pushback time; t_w is the actual wheels-off time; $\sigma(t)$ is standard deviation of the remaining taxi-out time at time t. In principle, variance in the remaining taxi-out time could be calculated at any time. However, it is more meaningful to discuss about the uncertainty at the critical stages. In this analysis, variance is calculated at 5 discrete stages during the taxi-out process: scheduled pushback, actual pushback, taxiway entry, queue entry and runway entry. With this, unpredictability could be further expressed as:

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

where, i is the index for the stage, 1 for scheduled pushback, 2 for actual pushback, 3 for taxiway entry, 4 for queue entry, 5 for runway entry and 6 for actual wheels-off time. For each scenario, $\bar{t}_i$ is the average time when stage i occurs, across all the aircraft. $\sigma(t_i)$ is the standard deviation of the remaining taxi-out time at stage i for all the aircraft too. Unpredictability based on this formula reflects the level of lack of predictability for each scenario overall rather than for individual flight.

2. Results on Unpredictability Comparison between Baseline and Advisory

Unpredictability is measured for all the 4 scenarios and the estimation results are shown in Fig. 2. We take the plot for the scenario with high traffic and advisory as an example to explain the results. In the plot, horizontal axis represents average time that has passed after scheduled pushback. 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, $i = 1$, standard deviation drops to about 2.3 min which is less than half of before. When the aircraft enters 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 min² which is the area below the blue step plot.
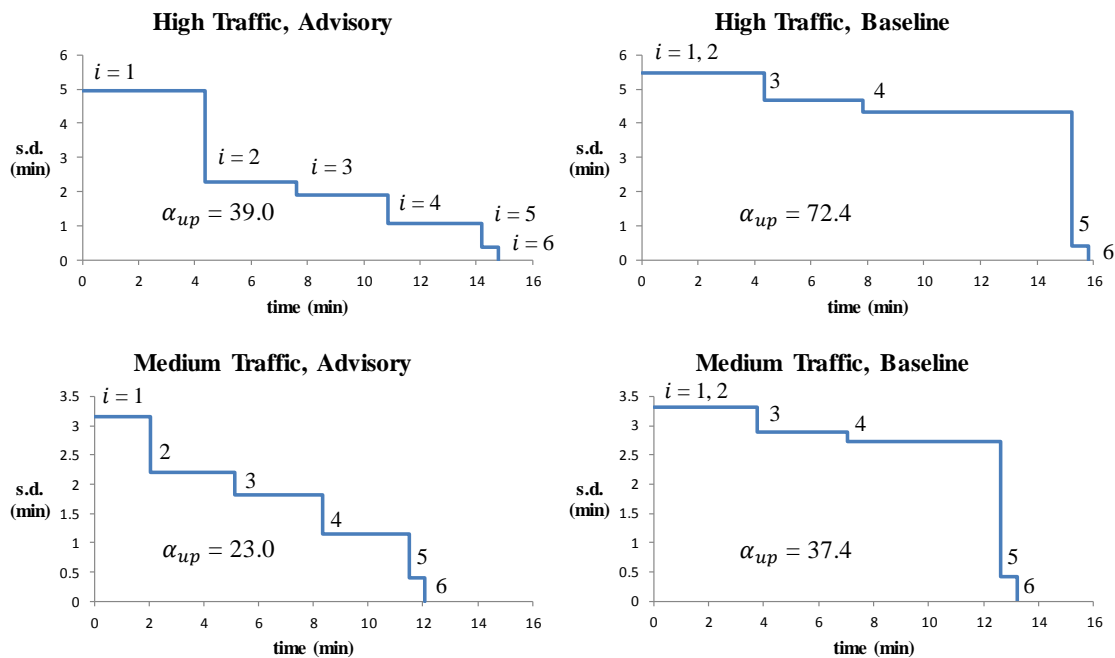

Figure 2. Unpredictability in total taxi-out time and its components (note that the y axis scales are different between high traffic and medium traffic)

American Institute of Aeronautics and Astronautics

At the same traffic density, the baseline scenario turns out to have much larger unpredictability, as shown in the plots on the right side. In total, taxi-out time in the baseline case is 1 minute more than that in the advisory case. However, the average time aircraft wait in the queue area is 7.4 min, more than double as it is in the advisory case. Pilots usually start the second engine when the aircraft is entering the queue area, and the engines provide the cabin power during the remaining taxi process. On the other hand, SARDA allows the aircraft to hold at the gate with the Auxiliary Power Unit (APU) providing cabin power when the aircraft is at the gate. A unit time of APU requires less fuel than a unit time of 2-engine on. Therefore, shifting delay from the queue area to the gate allows the aircraft to save fuel burn in the taxi-out process. Without advisories, variation in total taxi-out time is larger and does not reduce at the actual pushback because of absence of gate-holding. Moreover, uncertainty in taxi-out time is high until aircraft taxi into the runway where there is a big drop in the standard deviation. As a result, the area below the blue step plot is larger than that in the advisory case, which indicates less predictability. Similarly, in the medium traffic level total variance in the taxi-out time is lower with SARDA advisories and variance in taxi-out time is greatly reduced after actual pushback with the use of advisories.

In summary, uncertainty in wheels-off time declines faster with advisories at both traffic levels. Owing to the gate-holding, significant part of uncertainty in the taxi-out time is absorbed at the gate before pushback in the advisory case, especially for the high traffic scenario. In the baseline runs, actual pushback time of aircraft is about the same as the scheduled pushback time. Therefore, no reduction in variance is observed between these two times. After the aircraft pushback from the gate, uncertainty in the remaining taxi-out time also decreases at a faster rate for the advisory runs. At the time when the aircraft enters the queue area, majority of the uncertainty has been absorbed in the advisory runs whereas large uncertainty still exists in the taxi-out time in the baseline runs.

B. Regression analysis for Actual Taxi-out Time

Between actual pushback and takeoff, pilots need to complete multiple checklists, including After-Start-Checklist, Taxi-Checklist and Before-Takeoff-Checklist. Accurate predictions of taxi-out time will allow pilots to better plan the operations. Moreover, high predictability in takeoff time would enable pilots to start the second engine more efficiently. Considering this, linear regression analysis is performed to capture variables that may explain the variance in actual taxi-out time. To make the regression model predictive, only variables with values known at the time of pushback are employed in the regression analysis. For a given aircraft, taxi trajectory is well defined once the departure gate and the runway are given. This allows us to have a good estimate about taxi distance in each area on the airfield. In addition, we could easily count number of aircraft in each area before pushback.
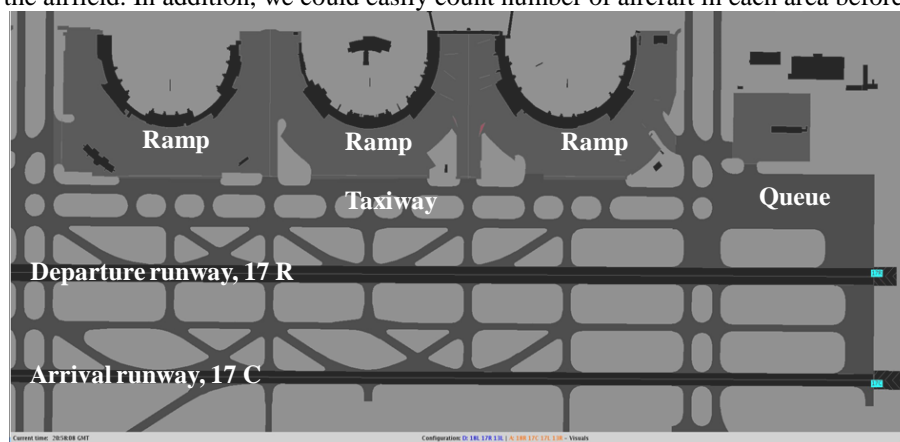

Figure 3. Runway Configuration Map

As shown in Fig. 3, for each departure aircraft, its taxi-out distance consists of four parts: ramp distance, taxiway distance, distance in the queue area and runway distance before wheels-off. For numbers of aircraft, we count numbers of aircraft at that time in the ramp, on the departure taxiway, in the queue area and in the arrival area. Arrival area is defined as the airfield below the departure runway in the map. Aircraft count in the arrival area is then all the arrival aircraft that have landed but not yet crossed the departure runway. The distance and count variables are selected as the independent variables in the regression model. We complete regression analysis for both traffic level scenarios with and without advisory. Here, we present estimation results for high traffic level scenarios only. Results for the medium traffic scenarios are similar, but with less difference between advisory and baseline cases.

5
American Institute of Aeronautics and Astronautics

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

Table 2. Estimation results on taxi-out time regression models, advisory and baseline

	Advisory		Baseline	
	Estimate	P Value	Estimate	P Value
Intercept	124.66	<0.0001	173.22	0.5235
Number of aircraft in the ramp	3.58	0.0195	26.12	<0.0001
Number of aircraft on the taxiway	5.77	<0.0001	29.67	<0.0001
Number of aircraft in the queue area	5.82	0.0008	12.56	0.0068
Number of aircraft in the arrival area	9.83	<0.0001	−	−
Ramp distance	0.38	<0.0001	0.95	<0.0001
Taxiway distance	0.15	<0.0001	0.05	<0.0001
Queue area distance	0.13	<0.0001	−	−
Number of observation	541		534	
Residual standard error	56.7		252.6	
Adjusted R^2	0.83		0.41	

When there are more aircraft along the taxi-out path, it is expected that taxi-out time will be longer. However, the incremental effect of each aircraft in the taxi-out path is larger in the baseline model, as shown by the larger coefficients. This shows SARDA mitigates the influence of other airfield traffic on the taxi-out time of a given aircraft. Number of aircraft in the arrival area is only significant in the advisory model. Therefore, marginal effect of arrivals is significant in advisory, but not in baseline. Coefficients for distances are positive for both models, with marginal effect of ramp distance the greatest. Without advisory, the queue area is congested with departure aircraft. The taxi time in queue area mainly depends on number of planes that will take off before a given aircraft rather than the queue distance of this aircraft. In general, coefficients for variables are larger in the baseline model. It reveals that there are fewer disturbances to a given aircraft from other operational parameters in the advisory model. With a larger adjusted R^2, advisory makes taxi-out time more explainable with the variables in the model. Standard deviation of the future response is expected to be much smaller with advisory, since residual standard error in the advisory model is only one fifth of that in the baseline model. If the regression models are used to predict the wheels-off time, the prediction in the advisory model would be much more accurate.

IV. Predictability from Traffic Management Perspective

Depending on the airport in consideration, departure runways usually experience queues and in some cases, specific runway queuing areas are present on the airport surface. With the assumption that pilots can identify their sequence of queue entry, if the takeoff sequence is well matched to the queue entry sequence, then the pilots can better estimate the takeoff sequence and time. Thus, from a traffic management perspective, it is important to ensure the sequence predictability in the queue area.

In normal runway operations at DFW, the departure queue has three lanes. This provides opportunities to the local controller to deviate from the First-In-First-Out (FIFO) structure of the queuing area for runway utilization maximization. Thus, in the baseline case, departure queue might not be served following FIFO.

The SARDA algorithms use multiple objectives (throughput and delay) to identify the takeoff sequence, which is periodically updated to address uncertainty in aircraft movement[3, 11]. However, there is no attempt to maintain the FIFO nature of the runway queue. The advisory would diminish the sequence predictability if the sequence of departure takeoffs provided by SARDA deviated from the potential FIFO sequence significantly. To assess the impact of SARDA on the performance, we compare the predictability performance in the advisory runs to that in the baseline runs.

Sequence predictability in each run is defined using the Spearman's coefficient[19], which measures the sequence correlation between the queue entry times and the aircraft wheels-off times. Mathematically, the metric can be written as:

$$\alpha_p = 1 - \frac{6\sum_{i=1}^{N}\left(R_i^{queue} - R_i^{roll}\right)^2}{N(N^2-1)} \tag{3}$$

6
American Institute of Aeronautics and Astronautics

where, N is number of aircraft in the run; R_i^{queue} is the rank of aircraft i in the queue entry time among all the aircraft; R_i^{roll} is the rank of aircraft i in the wheels-off time among all the aircraft. The value of the metric will be 1 if the rankings are in perfect agreement which indicates perfect FIFO performance and 0 if there is no relationship between the two rankings.

Using this formula, predictability is calculated for both the advisory and baseline runs at the two traffic levels: medium and high. The results are summarized in Table 3. The simulation runs are paired by controllers and the traffic scenario. The rank correlation is almost 1 in the advisory runs at both traffic levels. On average, rank correlation is more with advisories, which indicates higher sequence predictability. The improvement in predictability is more obvious at high traffic level. Taking a closer look at the detailed data, we find that one reason for a lower rank correlation in the baseline runs is the TMI flights. Without advisories, controllers tend to take 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 wheels-off time, and probably consumes more fuel.

Another reason for the better performance of the advisory 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 never exceeded 6 in the advisory case, whereas in the baseline case queue sizes of up to 12 were observed. Smaller overall number of aircraft in the runway queue potentially leads to less possibility of deviation from FIFO, and hence better sequence predictability in the advisory case.

Table 3. Results of Sequence Predictability in the Queue Area

Medium Traffic Level			High Traffic Level		
Runs	Advisory	Baseline	Runs	Advisory	Baseline
1	0.997	0.984	1	0.996	0.974
2	0.997	0.992	2	0.997	0.986
3	0.997	0.988	3	0.998	0.926
4	0.995	0.982	4	0.999	0.938
5	0.993	0.991	5	0.996	0.983
6	0.993	0.985	6	0.998	0.978
7	0.998	0.993	7	0.999	0.936
8	0.995	0.978	8	0.997	0.967
9	0.998	0.992	9	0.998	0.985
10	0.996	0.989	10	0.997	0.980
11	0.997	0.984	11	0.999	0.929
12	0.998	0.984	12	0.999	0.950
Average	0.996	0.987	Average	0.998	0.961
Test Statistics (t-test)	1.892		Test Statistics (t-test)	1.633	
Degree of Freedom	14		Degree of Freedom	11	

To test whether the improvement in predictability performance is statistically significant, we conducted a 1-tailed hypothesis test. The null hypothesis was that sequence predictability is the same with and without advisories, and the alternative hypothesis was that sequence predictability was higher in the advisory runs. Assuming unequal variances, the testing results reject the null hypothesis at a significance level 0.1 at both traffic levels. Therefore, improvement in sequence predictability in the queue area by SARDA is statistically significant.

Sequence predictability may also be assessed at a strategic level. Compared to the queue entry time, the scheduled pushback time of aircraft is known in advance. Assuming no congestion at the airport, we could estimate the scheduled wheels-off time for aircraft, which is calculated as the scheduled pushback time plus unimpeded taxi-out time. If the rank correlation between scheduled wheels-off time and real wheels-off time is strong enough, then the operations are predictable in terms of First-Scheduled-First-Served (FSFS). Here, we use scheduled wheels-off time instead of scheduled pushback time because the unimpeded taxi time for aircraft from different terminals can be very different. Since collaborative gate-holding is considered in the algorithm of SARDA, it is uncertain how the advisory affects the FSFS performance. To assess the impact of SARDA on the performance, the sequence predictability between the two wheels-off times is estimated following the same method as before. The results are summarized in Table 4.

American Institute of Aeronautics and Astronautics

Table 4. Results of Sequence Predictability between Scheduled and Real Wheels-off Time

Medium Traffic Level			High Traffic Level		
Runs	Advisory	Baseline	Runs	Advisory	Baseline
1	0.961	0.976	1	0.951	0.965
2	0.954	0.972	2	0.957	0.934
3	0.975	0.962	3	0.933	0.924
4	0.971	0.957	4	0.940	0.927
5	0.956	0.984	5	0.952	0.966
6	0.952	0.966	6	0.956	0.959
7	0.978	0.989	7	0.935	0.929
8	0.975	0.965	8	0.940	0.943
9	0.959	0.989	9	0.954	0.944
10	0.952	0.981	10	0.960	0.928
11	0.973	0.967	11	0.925	0.920
12	0.947	0.936	12	0.938	0.914
Average	0.963	0.970	Average	0.945	0.938
Test Statistics (t-test)	-0.423		Test Statistics (t-test)	0.365	
Degree of Freedom	22		Degree of Freedom	20	

Comparing the predictability of each run, there is no obvious trend in the difference in the performance. A 2-tailed hypothesis testing is conducted to test whether the predictability level is different for the two types of runs or not. With small test statistics, we accept the null hypothesis that the sequence predictability is the same with and without SARDA. In other words, there is no evidence to suggest that SARDA performs worse than baseline in maintaining FSFS operations.

V. Conclusions

In this paper, we define and measure predictability in airport surface operations from multiple perspectives based on NASA's experimental data from testing the SARDA, an airport surface management tool. Through a survey study, controllers report that SARDA reduces unexpected workload surges and facilitates handling of TMI flights. The survey results also show that the tool could potentially yield other benefits in terms of ability to handle off-nominal situations, avoid queue spillovers, etc. Controllers also indicated that the SARDA would be very beneficial when there is change in runway configuration or departure routes.

From operator's perspective, we study predictability in total taxi-out time, from scheduled pushback time to wheels-off time, considering two competing notions: timeliness of information and accuracy of information. Specifically, we define the counterpart, unpredictability, as the integrated standard deviation of the remaining taxi-out time over time from schedule pushback to actual wheels-off. The results show that SARDA reduces variability in total taxi-out time and eliminates uncertainty in taxi-out time sooner. This enables us to have a more accurate estimate of wheels-off time sooner and thus provides higher predictability in operations. Next, we conduct linear regression analysis of taxi-out time, from actual pushback to wheels-off. Taxi distances and numbers of aircraft in each area, such as ramp area and queue area, are selected to be the only explanatory variables because their values could be easily estimated at the pushback time and not depend on the following taxi-out process. The regression results indicate SARDA relieves traffic congestion on the airfield with less marginal effects of the variables. Moreover, aircraft wheels-off time could be predicted more accurately with SARDA giving a larger adjusted R^2 and a smaller residual standard error in the advisory model. Higher predictability from SARDA would allow pilots to better sequence their operations and start the second engine more efficiently.

Finally, we examine predictability in traffic management. Using sequence predictability, we assess the impact of SARDA on First-In-First-Out performance in the queue area and First-Scheduled-First-Served performance. The results show that SARDA improves the correlation between queue entry time and the real wheels-off time, which potentially allows pilots to make better estimates in wheels-off time. At a strategic level, although gate-holding is considered in the advisory runs there is no evidence to indicate that SARDA deteriorates FSFS performance.

Existing work on measuring predictability focuses on the variation in air traffic management system as experienced by the users. Recognizing the potential benefit from timeliness of information besides accuracy of information, predictability from flight operators' perspective is measured considering both aspects. Here, we assume predictability at a given time to be the product of the timeliness and accuracy. Different flight operators could have different sensitivities on timeliness of information according to their operational characteristics. Therefore, an

American Institute of Aeronautics and Astronautics

interesting extension of this work is to customize the relationship between the two features in predictability definition. More broadly, the relative importance of predictability, as compared to other dimensions of performance such as delay and efficiency, needs to investigated, both in the airport surface domain and in other parts of the NAS.

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-study 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)

Acknowledgments

References

[1] Hansen, M. "Micro-level analysis of airport delay externalities using deterministic queuing models: a case study," *Journal of Air Transport Management*, Vol. 8, 2002, pp. 73-87.

[2] Mueller, E. "Analysis of aircraft arrival and departure delay characteristics," *AIAA's Aircraft Technology, Integration, and Operations (ATIO)*, Los Angeles, California, 2002.

[3] Jung, Y., Hoang, T., Montoya, J., Gupta, G., Malik, W., Tobias, L., Wang, H. "Performance evaluation of a surface traffic management tool for Dallas/Fort Worth international airport," *9th USA/Europe Air Traffic Management R&D Seminar*, Berlin, Germany, June 14-17, 2011.

[4] Liu, Y., Hansen, M., Zou, B. "Aircraft gauge differences between the US and Europe and their operational implications," *Journal of Air Transport Management*, Vol. 29, 2013, pp. 1-10.

[5] Gilbo, E. "Airport capacity: representation, estimation, optimization," *IEEE Transactions on Control Systems Technology*, Vol. 1, No. 3, 1993, pp. 144-154.

[6] Ignaccolo, M. "A simulation model for airport capacity and delay analysis," Transportation Planning and Technology, Vol. 26, No. 2, 2003, pp. 135-170.

[7] Leigh Fisher, Landrum & Brown, George Mason University, University of California, Berkeley, Presentation & Design, Inc. "Evaluating airfield capacity," *Airport Cooperative Research Program Report 79*, 2012.

[8] Gupta, G., Malik, W., Jung, Y.C. "Effect of uncertainty on deterministic runway scheduling," *11th AIAA's Aircraft Technology, Integration, and Operations (ATIO)*, Virginia Beach, VA, September 20-22, 2011.

[9] Manzi, P., Lindberg, L.G., Wichman, K., Bleeker, O., Langhans, S. "Improving operational efficiency in the airport, airline and ATM processes," *Proceedings of Digital Avionics Systems Conference*, Salt Lake City, UT, October 24-28, 2004.

[10] Vormer, F., Marsden, A., Duong, V. "A complex systems approach to assessment of airport flexibility, predictability and efficiency," *27th International Congress of the Aeronautical Sciences*, Nice, France, September 19-24, 2010.

[11] Bradford, S., Knorr, D. Liang, D. "Performance measures for future architecture," *3rd USA/Europe Air Traffic Management R&D Seminar*, Naples, Italy, June 13-16, 2000.

[12] Air Traffic Services Performance Focus Group, CNS/ATM Focused Team. "Airline metric concepts for evaluating air traffic service performance," 1999.

[13] Gupta, G., Malik, W., Tobias, L., Jung, Y.C., Hoang, T., Hayashi, M. "Performance evaluation of individual aircraft based advisory concept for surface management," *10th USA/Europe Air Traffic Management R&D Seminar*, Chicago, Illinois, June 10-13, 2013.

[14] Malik, W., Gupta, G., Jung, Y.C. "Managing departure aircraft release for efficiency airport surface operations," *AIAA Guidance, Navigation, and Control Conference*, Toronto, Canada, August 2-5, 2010.

[15] Hoang, T., Jung, Y.C., Holbrook, J.B., Malik, W. "Tower controllers' assessment of the Spot and Runway Departure Advisor (SARDA) Concept," *9th USA/Europe Air Traffic Management Research and Development Seminar*, Berlin, Germany, June 14-17, 2011.

[16] Jung, Y.C., Hoang T., Montoya J., Gupta G., Malik W., Tobias, L. "A concept and implementation of optimized operations of airport surface traffic," *10th AIAA's Aircraft Technology, Integration, and Operations (ATIO)*, Fort Worth, Texas, September 1-15, 2010.

[17] Knorr, D., Post, J., Biros, J. "The operational assessment of free flight phase 1 ATM capabilities," *3rd USA/Europe Air Traffic Management R&D Seminar*, Naples, Italy, June 13-16, 2000.

[18] EUROCONTROL, Performance Review Commission. "An assessment of Air Traffic Management in Europe during the calendar year 2011," *Performance Review Report*, February, 2012.

[19] Blalock, H.M. *Social Statistics*, 2nd Edition. McGraw-Hill, Inc. USA, 1979.
.