2013-Gupta G-ATM2013-TN

Surface operations at airports in the US are based on tactical operations, where departure aircraft primarily queue up and wait at the departure runways. NASA's Spot And Runway Departure Advisor (SARDA) tool was developed to address these inefficiencies through Air Traffic Control Tower advisories. The SARDA system is being updated to include collaborative gate hold, either tactically or strategically. This paper presents the results of the human-in-the-loop evaluation of the tactical gate hold version of SARDA in a 360 degree simulated tower setting. The simulations were conducted for the east side of the Dallas/Fort Worth airport. The new system provides gate hold, ground controller and local controller advisories based on a single scheduler. Simulations were conducted with SARDA on and off, the off case reflecting current day operations with no gate hold. Scenarios based on medium (1.2x current levels) and heavy (1.5x current levels) traffic were explored. Data collected from the simulation was analyzed for runway usage, delay for departures and arrivals, and fuel consumption. Further, Traffic Management Initiatives were introduced for a subset of the aircraft. Results indicated that runway usage did not change with the use of SARDA, i.e., there was no loss in runway throughput as compared to baseline. Taxiing delay was significantly reduced with the use of advisory by 45% in medium scenarios and 60% in heavy. Arrival delay was unaffected by the use of advisory. Total fuel consumption was also reduced by 23% in medium traffic and 33% in heavy. TMI compliance appeared unaffected by the advisory

Tenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2013)

Performance Evaluation of Individual Aircraft Based Advisory Concept for Surface Management

Gautam Gupta, Waqar Malik, Leonard Tobias
University of California, Santa Cruz
NASA Ames Research Center
Moffett Field, California, USA

Yoon Jung, Ty Hoang, Miwa Hayashi
NASA Ames Research Center
Moffett Field, California, USA

Abstract— **Surface operations at airports in the US are based on tactical operations, where departure aircraft primarily queue up and wait at the departure runways. NASA's Spot And Runway Departure Advisor (SARDA) tool was developed to address these inefficiencies through Air Traffic Control Tower advisories. The SARDA system is being updated to include collaborative gate hold, either tactically or strategically. This paper presents the results of the human-in-the-loop evaluation of the tactical gate hold version of SARDA in a 360° simulated tower setting. The simulations were conducted for the east side of the Dallas/Fort Worth airport. The new system provides gate hold, Ground Controller and Local Controller advisories based on a single scheduler. Simulations were conducted with SARDA on and off, the off case reflecting current day operations with no gate hold. Scenarios based on medium (1.2x current levels) and heavy (1.5x current levels) traffic were explored. Data collected from the simulation were analyzed for runway usage, delay for departures and arrivals, and fuel consumption. Further, Traffic Management Initiatives were introduced for a subset of the aircraft. Results indicated that runway usage did not change with the use of SARDA, i.e., there was no loss in runway throughput as compared to baseline. Taxiing delay was significantly reduced with the use of advisory by 45% in medium scenarios and 60% in heavy. Observed gate-holds were less than 15 minutes in all but one scenario, and even in this scenario 95% of the aircraft had a gate hold of less than 15 minutes. Arrival delay was unaffected by the use of advisory. Total fuel consumption was also reduced by 23% in medium traffic and 33% in heavy. TMI compliance appeared unaffected by the advisory.**

Keywords; departure metering; airport surface traffic; integrated surface management; human-in-the-loop simulation

I. Introduction

Surface operations at airports in the US National Airspace System (NAS) are often based on tactical operations, with aircraft often being controlled reactively. Although there are variations in procedures at different airports, the basic procedures remain the same: departure aircraft are moved from gate to runway whereas arrival aircraft are moved from touchdown to the gate. There are numerous intermediate steps involved in these processes with some degree of connectivity in the steps. However, due to the mostly reactive nature of Air Traffic Control (ATC) at the airport and the lack of adequate decision support tools for surface management, surface operations can be inefficient. One cause of inefficiency is peak traffic: multiple aircraft pushback at around the same time and contest for the limited resource (the runway). This leads to many aircraft taxiing to the runway simultaneously causing long runway queues as well as adverse congestion effects on taxiways. Departure metering (holding the aircraft in a holding area, probably in the ramp) as a method to address this inefficiency has recently received attention in the literature. A recent Federal Aviation Administration (FAA) sponsored study estimated that at eight major US airports, departure metering could yield cumulative fuel savings of 2.3 billion USD from 2012 to 2030 [1]. Another study estimates potential fuel savings of 42 to 300 million USD in 2011 at 43 top US airports [2].

Given the potential benefits of departure metering, methods to implement it at US airports are being explored by federal organizations (FAA and NASA), academia and the aviation industry [3-6]. These concepts can be divided into two broad categories: Air Traffic Control Tower (ATCT) based and airline based. ATCT-based concepts provide "advisories" to ATCT controllers at the airport, specifically to the Ground Controller (broadly responsible for taxiway movements) and the Local Controller (broadly responsible for departure and arrival aircraft on runways). One such concept is based on controlling the rate of aircraft being released in to the taxiways by the ATCT controllers [6], as well as the number of aircraft in runway queues. In this rate control concept, rate advisories are provided to the Ground Controller only. Another concept, Spot And Runway Departure Advisor (SARDA), gives aircraft specific sequence and time advisories to ATCT to reduce the number of aircraft on the taxiways and runway queues [4]. In the SARDA concept, advisories are provided to both Ground and Local Controllers. Both the rate control and SARDA concepts are based on moving the delay from the taxiways and runway queues to the ramp area by providing guidance to the ATCT, and do not cause a reduction in overall delay. Further, they are both "tactical" tools, which respond to the current traffic scenarios with little strategic planning.

Another set of concepts reduce the number of aircraft in taxiways and runway queues by altering the push-back times for departure aircraft; these alterations are done in collaboration with the airlines. One such method, called Collaborative Departure Queue Management (CDQM), manages the length of the runway departure queues by giving the flight operator an allocation of slots to enter the taxiways [3]. The assignment of the slots is done through a "ration by schedule" approach, which effectively is a first-scheduled-first-serve approach. Another method with some similarities to CDQM was developed and implemented at the John F. Kennedy Airport in

New York, and is currently under use [5]. The slot allocation is conducted using ration by schedule, and slots are assigned two hours in advance. Any swaps or changes in this allocation are managed by the "slot allocation manager," a neutral third party. The idea behind both these approaches was to hold aircraft at the gate or pre-assigned holding pads with engines off as much as possible, reducing the delays on the taxiway as well as the fuel consumption and emissions. Compared to the above ATCT advisory-based tools, these airline Collaborative Decision Making (CDM) tools are strategic in nature.

Tools for departure management have been studied in the European airports' context even before the studies in US. The Eurocontrol and DLR DMAN (Departure MANager) tool [7] does tactical departure management through a heuristics based optimization scheme. The tool provides a "managed time" for pushback, and initial results showed no change in runway throughput while reducing taxi-out times. However, in [7] it is emphasized that DMAN does not use surveillance information and hence could provide target runway sequences that are not achievable due to interactions between aircraft on the taxiways. These interactions have been the focus of recent research [8, 9] where the DMAN tool is being integrated with SMAN (Surface MANager); the position of each aircraft is taken into account (through surveillance data) and the runway schedule is updated. This frequent updating becomes even more important at US airports, where "with the exception of taxi-in times, variability in times for all flight phases is higher" as compared to airports in Europe [10].

Recently, a new version of the SARDA system was introduced [11] that combined the tactical controller advisories with collaborative gate hold to develop an integrated system for airport surface management. As before, this system is *aircraft specific*. It is capable of collaborative gate holding either strategically (where pushback times are provided an hour before scheduled pushback time) or tactically (where pushback hold is recommended, if necessary, only when the aircraft is ready to pushback). In [11], the tactical gate hold version of the new SARDA system was tested in an automated simulation setup with varying levels of uncertainty in gate pushback.

In this paper, the results from a human-in-the-loop (HITL) evaluation of the tactical gate hold version of the new SARDA system are presented. The HITL study was conducted at the NASA Ames Research Center in May 2012 with recently retired controllers and pseudo pilots in a 360° simulated tower environment. There were numerous improvements in both the SARDA concept and experiment setup as compared to the previous tests [4]:

- Holding at gate was added as compared to the holding at spot.
- The SARDA scheduler was updated to a unified single scheduler providing gate hold, Ground Controller and Local Controller advisories.
- Taxi movements were more realistic due to taxi speed uncertainty.
- An Electronic Flight Strip (EFS) based interface was designed for the ATC controllers, and SARDA advisories were incorporated in EFS.
- The effect of Traffic Management Initiatives (TMI) on certain aircraft was evaluated. TMI is used here as a generic term for extra timing constraints on certain aircraft due to weather or other phenomenon, is described further in section IV.
- A complete out-of-the-window view was provided to ATC controllers in the simulated tower environment.

A variety of data was collected in this study, and this paper presents some of the system performance data from the experiments. In the rest of the paper, first the tactical gate hold SARDA concept is described. Then the experimental setup is described. This is followed by system performance results, and a short summary of effect on controller workload. The paper concludes with a discussion and directions for future work.

II. SARDA Concept Description

Previous work on SARDA was on developing Ground and Local Controller advisories for improved surface traffic management [4]. The new version of SARDA under development is the augmentation of Ground and Local Controller advisories through sharing of flight movement and related operations information between airport operators, flight operators and ATC. The goal is to improve airport surface operations by maximizing available airport and airspace capacity while minimizing adverse effects on stakeholders, passengers and the environment. In previous work, the taxi and runway-queue delay for departure aircraft was moved to the spot (the transition point where ATC tower takes control of the aircraft from ramp controller). Here, the delays are moved all the way to the gate to provide benefits to the airlines in fuel savings and potentially better connection time for passengers. The concept includes sharing data between ATCT and airline operators (including updated gate pushback readiness and ATC constraints due to weather), which enables these benefits.

Some of the key assumptions for the SARDA concept are:

- Ramp area operations are managed by airlines or airport authorities and, therefore, ATCT does not have direct control of gate push back for departure aircraft.
- Voice is still the main means of communication between ATCT controllers and pilots.
- Aircraft position data may not be available in the ramp area through a surface surveillance system, such as Airport Surveillance Detection Equipment – Model X (ASDE-X). However, the actual gate push-back time for departure aircraft is known. Aircraft position data at the spots and the movement area is available through ASDE-X, with the assumption that the current accuracy levels in ASDE-X would be sufficient.
- Prediction of arrival times of departure aircraft at runway queue entrance is available. In the current implementation, predictions based on nominal unimpeded travel from scheduled gate push-back to the runway are used. This could be changed in the future, and better prediction models can be used.
- For the cases where arrival aircraft cross the departure runway, prediction of earliest runway crossing time is available. These can be inputs from a system like Traffic Management Advisor (TMA).

It should be noted that the version of SARDA being discussed here includes tactical gate hold. For strategic gate hold, a different version of SARDA has been described in [11].

A. SARDA Scheduler

The core computational engine behind SARDA, called the SARDA scheduler is based on the spot release planner (SRP) [12], a method to provide metering advisories. SRP is a two-stage algorithm: The first stage is a runway scheduler [13, 14], which gives the best sequence and times for runway usage by a set of departure aircraft ready for take-off and arrival aircraft waiting to cross the same departure runway. The second stage of the SRP determines times to release aircraft from gates or assigned spots to meet the optimal departure schedules. It should be noted that the implementation of the SARDA scheduler is airport specific. Some airports may have multiple runways with some degree of inter-dependence in operations, some airports might have multiple runways which can operate independently and others might have separate arrival and departure runways. The runway scheduler in the first stage of SRP is tailored to the runway layout as well as the runway configuration. The second stage of the SRP is tailored to the airports' taxiway layout. Where applicable, the second stage of SRP can also provide taxi route advisories.

The scheduler takes as input the current snapshot of the airport, aircraft specific parameters, separation constraints, scheduled push-back times and scheduled arrival times for the aircraft in the next 15 minutes. Uncertainties in aircraft movement pose a challenge to generating advisories. To mitigate the effect of uncertainties, the scheduler gets an updated airport condition snapshot every 10 seconds, which is then used to recalculate the schedule. Thus, the scheduler follows an update cycle of 10 seconds. In order to reduce the frequent changes in advisories for the controller due to the frequent update, a "freeze horizon" is implemented: the schedule for certain aircraft does not change based on the location of aircraft and the current time. For example: in the Local Controller advisory, the sequence for the first three aircraft scheduled to use the runway is fixed between successive scheduled calls.

As mentioned before, the SRP's first stage is a runway scheduler that provides the best runway usage schedule for both arrivals (crossings) and departures. The runway scheduling problem has numerous constraints (wake vortex separation, miles-in-trail and others) and can be solved for multiple objectives including throughput (runway usage time for last aircraft) and system delay (total delay for all aircraft). Reference [13] shows that optimizing for system delay results in small deviations from optimal throughput, whereas optimizing for throughput results in large deviations in system delay. For this reason, system delay was chosen as the objective for the scheduler. Further, taxi time estimates were required for the prediction of earliest take-off times for the first stage of SRP, and for calculating spot and gate releases in the second stage. These predictions were based on the historical aircraft movement on the surface: numerous unimpeded trajectories on surface were analyzed to estimate the distribution of unimpeded speeds. The speed corresponding to given percentile was chosen for the estimates in the first stage of SRP, and for the second stage as well. Appropriate selection of this percentile provides the buffer in the runway queue to ensure that the runway is not under-utilized. For these experiments, the 30th percentile of unimpeded speed was chosen for both the stages of SRP.

B. SARDA Tactical Gate Hold Walkthrough

Below, the SARDA tactical gate hold concept is illustrated using an example of flight XYZ101 with scheduled pushback time of 10:00am.

- 10:00am Scheduled pushback time for the aircraft. Due to ongoing boarding or baggage loading, the pilot is not ready to pushback
- 10:03am Pilot communicates pushback readiness. Ramp controller inputs readiness in the ramp controller interface. Through collaboration with ATC, this interface reflects any TMI constraints that aircraft might have. Based on current traffic, SARDA modifies actual pushback to 10:12am. Ramp controller accepts this and communicates this to the pilot.
- 10:12am Ramp controller interface prompts for XYZ101 pushback. Ramp controller communicates this to the pilot. Pilot accepts and commences pushback.
- 10:14am Pilot reaches the spot to transition to the active movement area. Ground Controller is aware of XYZ101, but SARDA advises spot release close to 10:15am due to taxiing traffic next to the spot.
- 10:15am Ground Controller initiates communication with XYZ101 and clears it to taxi using one of the standard taxi routes. Pilot acknowledges, and commences taxiing.
- 10:18am Pilot reaches departure queue. Local Controller is aware of XYZ101. SARDA advises aircraft is 3rd in the departure queue. Local Controller can communicate this to the pilot at his discretion
- 10:20am Local Controller clears XYZ101 to proceed to the runway, line up and wait.
- 10:21am Local Controller clears XYZ101 for takeoff. Pilot acknowledges and begins takeoff procedure.

III. Experimental Setup

This section describes the setup for the HITL experiments to evaluate the tactical gate hold version of SARDA. The experiments were conducted at the NASA Ames Research Center in the FutureFlight Central (FFC) facility, which offers a 360-degree, full-scale, real time simulation of an airport control tower. Multiple software systems were integrated at FFC to enable pseudo-pilot control of aircraft, ATCT controller displays and Computer Generated Imagery (CGI) based out-of-the-window view from the ATC tower. In the following subsections, the system architecture, controller user interface and experiment scenarios are discussed.

An underlying assumption for these experiments was 100% airline compliance with SARDA generated gate pushback times. It was assumed that pilots would call in pushback readiness at scheduled pushback times; SARDA would then generate pushback times to which pilots would comply. The effects of various levels of airline compliance with SARDA pushback times has been studied in an automated setup before [11], but due to certain system limitations as well as ongoing development of the ramp controller interface, uncertainty in meeting SARDA generated pushback times was not tested.

A. System Architecture

The tactical gate-hold version of SARDA was implemented within the Surface Management System (SMS) [15]. SMS was originally developed as a decision support tool to assist ATCT controllers and managers as well as airline operators in managing and controlling airport surface operations [16]. For this simulation, SMS exchanged flight information and scheduling solutions with the optimization algorithms over the network. Existing SMS user interfaces were modified to provide advisories to the Ground and Local Controller positions.

The Airspace Traffic Generator (ATG) system was used to generate motions of aircraft either on the surface or in the airspace near the airport, and sent position data to SMS for display [17]. ATG is a high-fidelity, real-time aircraft simulation tool that provides the capability to move the aircraft on the airport surface and generate and display targets of the aircraft. The Ground Pilot Stations (GPSs), components of ATG, were used by the pseudo-pilots to manually taxi aircraft when clearances were issued by the controllers over voice.

B. Experiment Scenarios

The experiments were conducted using the east side of the Dallas/Fort Worth International Airport (DFW), with FFC providing the view from the east DFW ATC tower. Figure 1 gives the map view of the east side of DFW with various aspects marked.

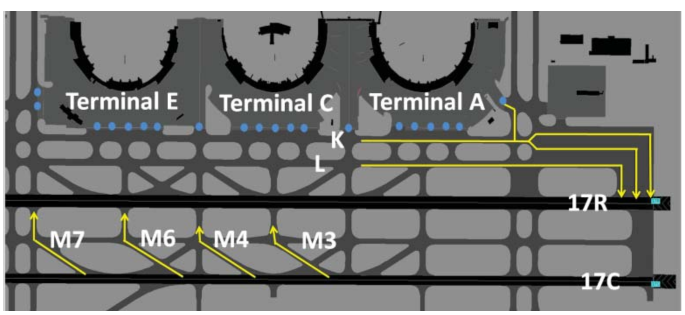

Figure 1: East side of DFW showing departure runway (17R), arrival runway (17C), arrival exits and departure taxi routes.

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). These scenarios were labeled Medium 1, Medium 2, Heavy 3 and Heavy 4. The medium scenarios had 40 departure aircraft in 50 minutes, and heavy had 50 departure aircraft in 50 minutes. Figure 2 shows the cumulative departure counts for the four scenarios at 2 minute intervals. These were plotted assuming unimpeded travel from scheduled push-back. It shows that Heavy 3 has 2 peaks between the 10^{th} and 30^{th} minute, whereas Heavy 4 has only one large peak starting at the 25^{th} minute.

Scenarios were generated such that departure traffic begins at the gates upon activation in the simulation and arrival aircraft appear about 10 nmi from the runway threshold. After a departure aircraft pushes back from the gate, it maneuvers in an automated mode towards its assigned spot and stops before it, unless the Ground Controller issues the pilot a taxi clearance while the aircraft is still moving. Scenario data for a departure aircraft contain callsign, aircraft type, flight plan route, departure fix, activation time (push back or first track hit), initial position, gate, spot, and runway. Uncertainty in taxiway movement was included, with each aircraft having a nominal taxi speed between 12 to 17 knots. Of course, the aircraft slowed down when instructed by the pseudo pilot; the above nominal speed is the default taxi speed for the aircraft unless instructed otherwise by the pseudo pilot. For repeatability, the same aircraft was given the same nominal speed in multiple runs of the same scenario.

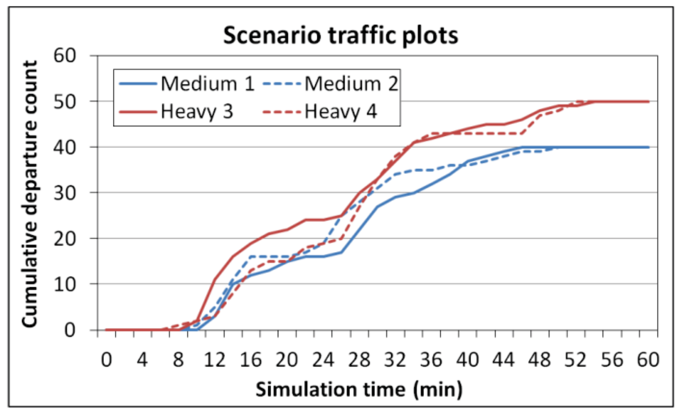

Figure 2: Cumulative departure aircraft counts for the four experiment scenarios

To test the effect of the SARDA advisories, a non-SARDA or baseline case was also tested. In this case, the SARDA scheduler was not running, and advisories were absent with no gate holding. This essentially represents current day operations. To determine the performance envelope of the concept, controllers were asked to follow the advisories, when present, as long as it did not compromise the safety of the aircraft.

Six recently retired controllers participated in the experiments over a period of three weeks, with each pair running scenarios for an entire week. Controllers rotated between the Ground Controller and Local Controller position, with a controller running each scenario twice at the same position, once in baseline and once in SARDA or advisory case. This results in a total of 48 runs, with six runs of each scenario in either baseline or advisory mode.

C. Controller Interface

Displays for both Ground and Local Controllers were provided for the HITL simulation. The basic display for the Ground Controller is composed of an existing SMS map display that shows spots and taxiways under the controller's responsibility. The basic display for the Local Controller consists of a surface map of the responsible area (i.e., runway queue and crossing queues) and a map of terminal airspace that covers portions of final approach and initial climb paths. However, instead of using paper strips, an Electronic Flight Strip (EFS) system was developed for both the advisory and baseline case. The EFS system is touch screen enabled, and mimicked certain functionalities of the paper strips. In the advisory case, the controller advisories were integrated in the EFS system and displayed to the controller. The Ground Controller was shown the spot release sequence and window; the Local Controller was shown the sequence of runway 17R operations including departure takeoffs and arrival crossings. Further, advisories on departure taxi route were also generated and presented to the controller within the EFS system. Taxiing and airborne aircraft are shown on the map displays with a data tag attached to the aircraft icon. Figure 3 shows a sample picture of the Ground Controller display with both the map and EFS. A detailed description of the controller interface can be found in [18].

Figure 3: Map (left) and EFS (right) displays for the Ground Controller

IV. RESULTS

This section details the results observed during the HITL experiments. The results are split into runway usage, delays, fuel and effect on TMI aircraft. The section concludes with a brief note on controller workload.

A. Runway Usage

Runway usage is defined as the number of runway operations in a given time period. Operation would include departure take-off or arrival crossing on runway 17R. With the SARDA concept enabling departure metering at the gates, it is important to compare the runway usage to detect cases of excessive metering. Figure 4 shows the cumulative runway usage for the four scenarios at five-minute intervals. Each bar represents the mean over the six runs of the scenario in either baseline or advisory case, and the whiskers represent the max and min over the six data points.

The results show almost no difference in throughput between the advisory and baseline runs of the same scenario. Hence, no loss in runway usage was observed for the SARDA runs with departure metering.

B. Delays

Delay is defined as the difference between the observed travel time and the unimpeded travel time for the same route or travel section. Since the nominal speed of the aircraft could vary from 12 to 17 knots (section III.B), an unimpeded speed of 17 knots is used for calculating unimpeded travel times to avoid negative delay values. First, the delay in departure aircraft is considered, and then the arrival delay.

1) Departure Aircraft

For departure aircraft, first the scheduled delay is considered. Scheduled delay is defined as the observed take-off

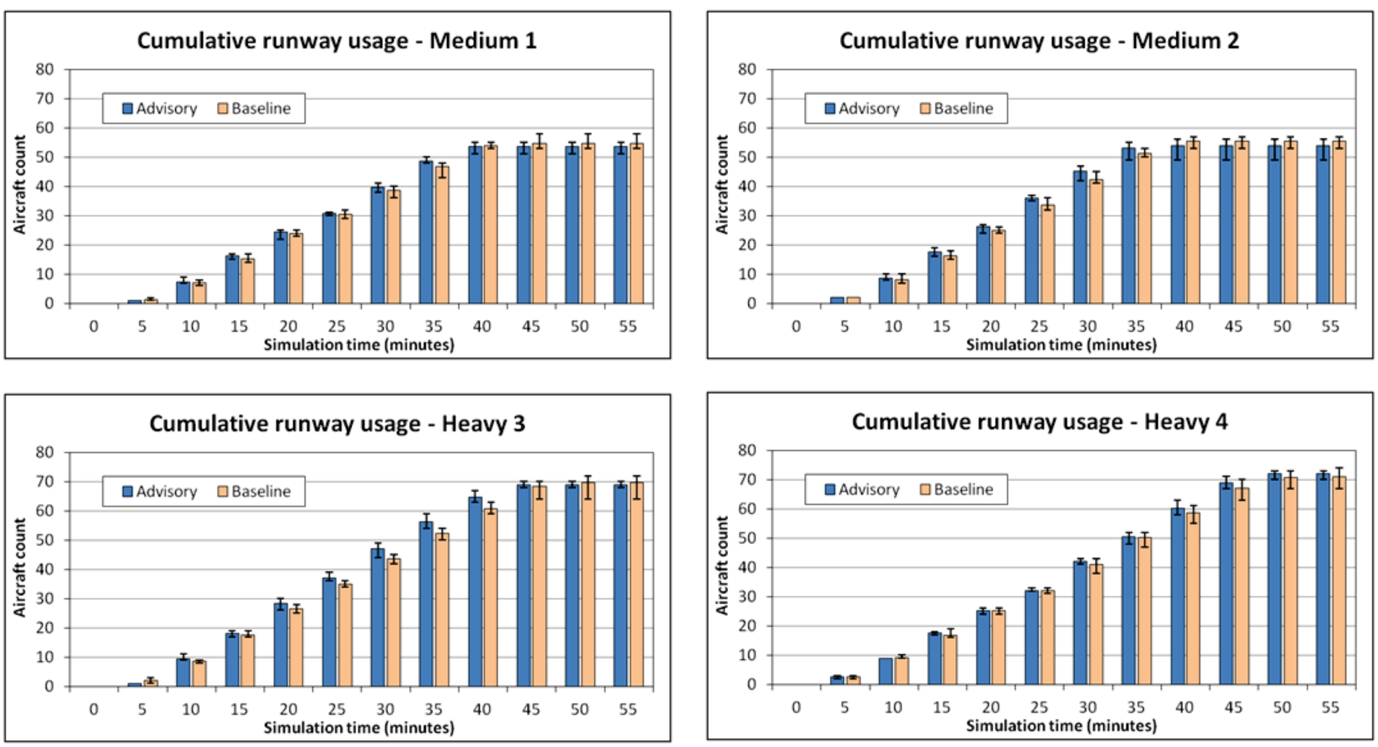

Figure 4: Throughput comparison in advisory and baseline runs for all scenarios (bar represents the mean; whiskers represent the min and max)

time minus the unimpeded take-off time calculated from the scheduled pushback time. Scheduled delay is the overall delay experienced by the aircraft at gate, ramp, taxiways and runway queue. Figure 5 shows the box and whisker plots for both advisory and baseline runs for the four scenarios. As expected, scheduled delay increases with increasing traffic volume. It appears that scheduled delay is slightly less with the use of advisory; this effect is further discussed in subsection E.

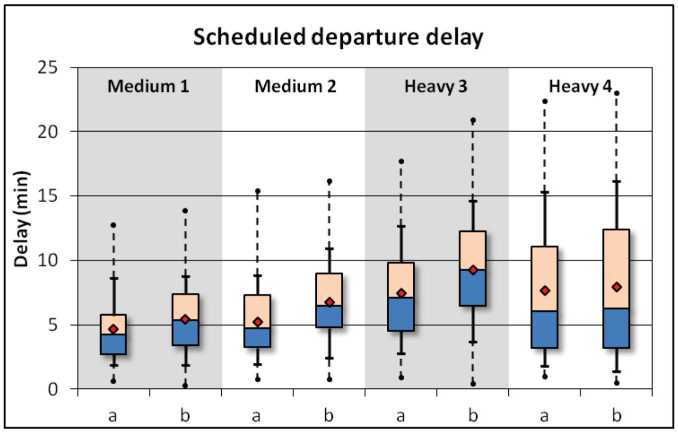

Figure 5: Scheduled departure delay. Horizontal axis "a" and "b" represent advisory and baseline. Bars show median, 25th and 75th percentile; whiskers show 10th and 90th percentile; dots show min and max; diamonds show mean.

One of the prime motivations for departure metering is to move the delay from the taxiways and runway queues to the gate, to enable less congestion in the active movement area and fuel savings. To gauge this effect, taxiing delay is evaluated, and is defined as the difference in observed and unimpeded take-off times, when the unimpeded take-off has been calculated from the actual push-back time. This represents the delay in the ramp, taxiways and runway queues. For baseline runs, this is equivalent to the schedule delay, but is less for the advisory runs. Figure 6 presents the taxiing delay results. As expected, taxiing delay is noticeably reduced with the use of advisory: an average 45% (3 min per aircraft) reduction in medium traffic and 60% (5.5 min per aircraft) reduction in heavy traffic were observed. Further, the variation in taxiing delay increases with increasing traffic in the baseline cases, but is insensitive in the advisory case.

The current FAA on-time performance metrics[1] would preclude gate-holds of more than 15 minutes beyond the scheduled pushback time, since these would be reported as airlines delay. Although a policy change can be implemented in the future after discussions with various stakeholders, an immediate solution would be to limit gate hold. The SARDA system is capable of limiting gate hold, but this functionality was not implemented in the experiments. Thus, it becomes important to investigate the gate hold on each aircraft. Figure 7 shows the gate hold values across all the experimental runs. It should be noted that even without limiting the amount of gate-hold, in all scenarios except Heavy 4, gate hold was never more than 15 minutes. Even in Heavy 4, 95% of the aircraft had a gate hold of less than 15 minutes.

[1] Code of Federal Regulations (CFR) Title 14-Aeronautics and Space, Chapter II, Part 234 – Airline Service Quality Performance Reports

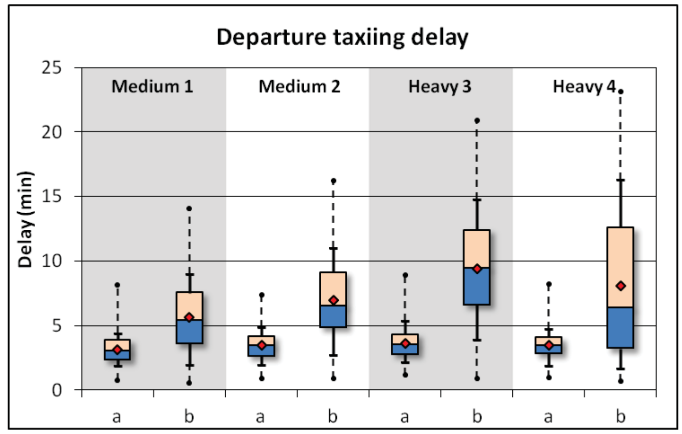

Figure 6: Taxiing delay for departures. Horizontal axis "a" and "b" represent advisory and baseline. Bars show median, 25th and 75th percentile; whiskers show 10th and 90th percentile; dots show min and max; diamonds show mean.

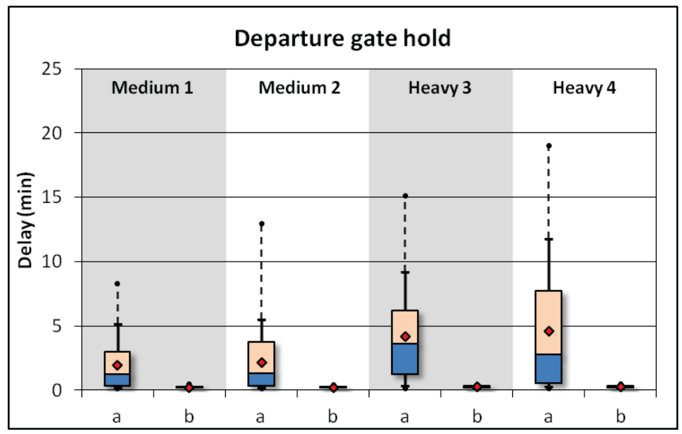

Figure 7: Gate hold for departures. Horizontal axis "a" and "b" represent advisory and baseline. Bars show median, 25th and 75th percentile; whiskers show 10th and 90th percentile; dots show min and max; diamonds show mean.

2) Arrival Aircraft

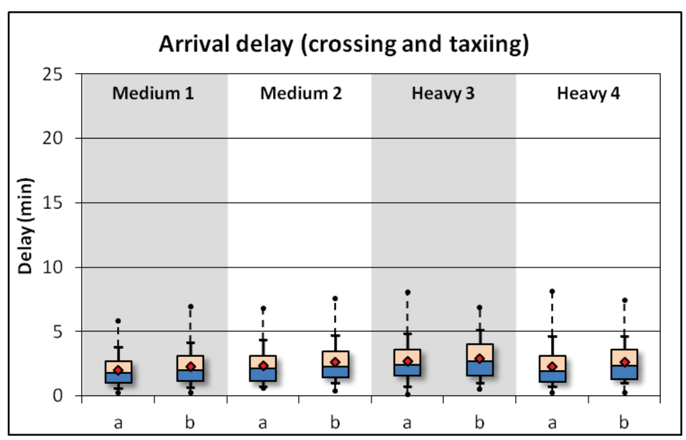

Figure 8: Arrival delay. Horizontal axis "a" and "b" represent advisory and baseline. Bars show median, 25th and 75th percentile; whiskers show 10th and 90th percentile; dots show min and max; diamonds show mean.

The delay in arrival aircraft is investigated to check if the use of advisory has an effect on arrival aircraft. Since arrival-crossing advisories are being provided to the Local Controller, it is a reasonable concern whether the improvements in departure taxiing delay are being achieved by increasing arrival delays. Figure 8 shows the delay in arrival aircraft across all runs; this includes delay in waiting to cross as well as taxiing to the spot after crossing. Arrival taxiing in the ramp is not considered here, since this was handled automatically by the ATG software. Results show that the delay in arrival aircraft were insensitive to the use of advisory or traffic level.

C. Fuel Consumption and Emissions by Departure Aircraft

The fuel consumed by departure aircraft was evaluated using the method described in [19]. This method is an augmentation of the current ICAO emissions databank, and includes effects of stops and acceleration events. For departure aircraft, fuel consumption was evaluated assuming all-engine taxiing from actual pushback. It was assumed that if the aircraft is being held at the gate, the engines are off. Figure 9 shows the total fuel used by departure aircraft over all the experimental runs. Fuel consumption is smaller in the advisory case as compared to the baseline: 23% average reduction in medium traffic and 33% average reduction in heavy traffic was observed. Moreover, fuel consumption in the baseline case seems more sensitive to traffic level as compared to advisory.

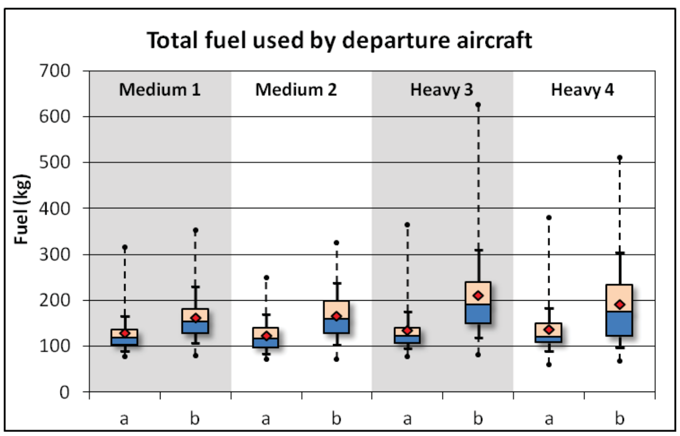

Figure 9: Total fuel in departure aircraft. Horizontal axis "a" and "b" represent advisory and baseline. Bars show median, 25th and 75th percentile; whiskers show 10th and 90th percentile; dots show min and max; diamonds show mean.

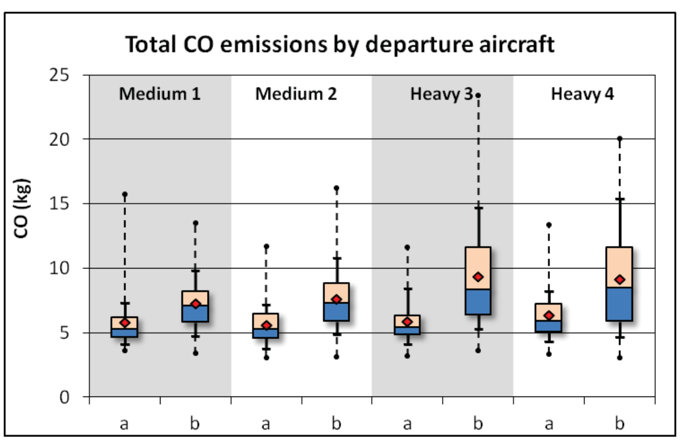

Figure 10: Total CO emissions by departure aircraft. Horizontal axis "a" and "b" represent advisory and baseline. Bars show median, 25th and 75th percentile; whiskers show 10th and 90th percentile; dots show min and max; diamonds show mean.

Along with fuel consumption, emissions from departure aircraft were also evaluated. Emission estimates are generated using emission indices, which are the amount of emissions generated per kilogram of fuel at certain engine thrust levels [19]. Hence, the trends in emissions are similar to those in fuel consumption. Due to restricted space, only carbon monoxide (CO) emissions are provided here; hydrocarbon (HC) and nitrogen oxides (NOx) show similar trends. Figure 10 shows CO emissions calculated for all departure aircraft in the experimental runs. As in fuel consumption, CO emissions are lower for the advisory case with a 24% average reduction in medium traffic and 34% average reduction in heavy traffic.

D. Effect on TMI Aircraft

As discussed before, certain aircraft in each scenario had TMI's associated with them; these were the desired takeoff times given to a subset of departure aircraft. Five aircraft had TMI's in the medium traffic scenarios, and seven aircraft had TMI's in the heavy scenarios. The window specified to the controllers for TMI adherence was ± 60 seconds. If the controllers were somehow unable to meet the TMI time, they were asked to get the take-off close to the TMI time, and no new TMI was generated.

In this section, the controllers' adherence to the prescribed TMI times is explored. For each aircraft, the observed take-off time was compared to the prescribed time. The observed takeoff time is defined as the time the departure aircraft crosses a threshold on the runway, close to taxiway Z. There is a small delay between the controller command for takeoff and the time the aircraft crosses this threshold; this could be due to variation in pilot response. Thus, when considering adherence to TMI time, a strict ± 60 second window would be incorrect. For this purpose, the actual difference in takeoff and prescribed time is considered. Figure 11 shows the plot of this difference for the various runs.

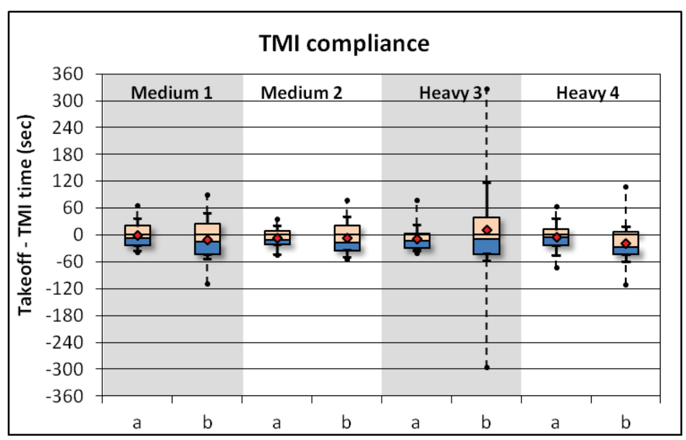

Figure 11: Difference in takeoff and TMI time. Horizontal axis "a" and "b" represent advisory and baseline. Bars show median, 25th and 75th percentile; whiskers show 10th and 90th percentile; dots show min and max; diamonds show mean.

In almost all cases, 90% of the TMI aircraft seem to meet the prescribed window. Besides the few outliers in the heavy scenario, there does not seem to be any evidence that TMI compliance was affected by the use of the advisory. However, it was expected that the primary effect of the advisory on TMI aircraft would be a reduction in delay. It is possible that to reduce the possibility of missing a TMI time, controllers might get the TMI aircraft to the runway sooner and then make it wait in the departure queue. With the use of the advisory, it was expected that such cases would decrease. This is explored further in the next section.

E. Statistical Significance of Taxiing Delay Reduction

Figure 6 shows that the use of advisory decreases the taxiing delay, and the effect is more prominent for the heavy traffic cases. In order to parse and understand such effects, tests of statistical significance need to be done.

The first step in testing the statistical significance of the benefits of advisory and the effect of traffic level on taxiing delay would be a t-test. However, a requirement for the use of t-test is the independence within the sample, i.e. there should be no dependence within the dataset. For taxiing delay, it is highly probable that part of the delay from the leading aircraft in the runway takeoff sequence might propagate to the trailing aircraft. Thus, the delay of the n^{th} aircraft in the runway takeoff sequence might depend on the delay of the $(n\text{-}1)^{th}$ aircraft. To address this effect while conducting tests for significance, linear regression is used with the delay of the previous departure aircraft as a variable within regression.

It should be noted that the using the delay of the previous aircraft as an independent variables is a case of the lagged variable model where disturbances are serially correlated [20]. Ordinary least squares regression would not give unbiased estimates of the coefficients in this case. To address this, a first order autoregressive process was assumed and generalized least squares was used, with estimates evaluated using maximum likelihood estimation.

Besides the use of advisory and the traffic level, the effect of aircraft weight class and an aircraft having a TMI was also tested. Further, it is possible that a longer taxi distance will lead to larger delays, since it increases the possibility of encountering other aircraft while taxiing. For this reason, total taxi distance was also included as an explanatory variable. Below is a list of all the variable used in the regression model, and the description of the variables. Along with these variables, the interaction terms with the indicator variables are also explored.

A	Advisory or baseline, 1 if advisory, 0 otherwise
H	Traffic scenario, 1 if heavy, 0 otherwise
W	Aircraft weight class, 0 if large, 1 otherwise
TMI	1 if aircraft had TMI
L	Distance traveled in ramp, taxi and queues
D_{n-1}	Taxiing delay for the immediate predecessor in runway sequence (departure aircraft only)

Table 1 gives the results from the linear regression. The estimate, standard error, t-value and significance level are presented; rows in white are significant to the 5% level, rows in grey are not. Following are some of the observations from the regression:

- The effect of advisory is significant, and the use of SARDA advisory decreases taxiing delay in all cases.
- Delays are higher in heavy traffic. However, the use of advisory reduces delay in heavy traffic also. The magnitude of the interaction term A×H is close to that of H; this suggests that the use of advisory substantially reduces the increased delay effect of heavy traffic.
- TMI aircraft typically have higher delay than non-TMI aircraft, and this effect is significant. However, the use of advisory seems to reduce this effect (A×TMI), and the reduction is statistically significant. Further, heavy traffic scenarios further increase the delay in TMI aircraft. A potential reason for this is the presence of three runway queues at DFW. Controllers use one of the queues ("full length") to stage TMI aircraft, a procedure that allows for a TMI departure at anytime but increases the possibility of more TMI delays in runway queue.
- Aircraft in heavy weight class seem to get larger delays in both advisory and baseline case, although the effect is borderline significant.
- Delay does not depend on the taxi distance in baseline case, but there is some correlation in the advisory case (A×L).
- The delay experienced by the previous aircraft in runway sequence significantly affects the delay of the successor. This is true for both baseline and advisory, but magnitude of the effect is reduced in the advisory case (A×D_{n-1}). The effect seems to reduce in the heavy scenarios, but this is not statistically significant.

TABLE 1: RESULTS FROM LINEAR REGRESSION ON TAXIING DELAY

Variable	Coefficient Estimate	Std. error	t-value	Pr(>|t|)
Intercept	154.70	30.44	5.08	0.00
A	-159.82	31.49	-5.08	0.00
H	104.58	32.88	3.18	0.00
W	86.69	46.39	1.87	0.06
TMI	309.43	53.19	5.82	0.00
L	-0.01	0.01	-0.51	0.61
D_{n-1}	0.65	0.04	16.58	0.00
A×H	-74.93	16.07	-4.66	0.00
A×W	31.72	24.62	1.29	0.20
A×TMI	-160.09	24.50	-6.53	0.00
A×D_{n-1}	-0.23	0.07	-3.27	0.00
A×L	0.06	0.01	5.42	0.00
H×W	-5.99	22.73	-0.26	0.79
H×TMI	84.59	21.88	3.87	0.00
H×L	-0.01	0.01	-0.81	0.42
H×D_{n-1}	-0.07	0.04	-1.51	0.13
W×L	-0.03	0.01	-1.76	0.08
W×D_{n-1}	-0.23	0.05	-4.45	0.00
TMI×L	-0.06	0.02	-3.55	0.00
TMI×D_{n-1}	-0.30	0.06	-5.25	0.00

F. Statistical Significance of Scheduled Delay Reduction

Figure 5 hints at some reduction in scheduled delay. To explore this further, a similar linear regression treatment of scheduled delay was conducted, as done in the previous section. All the same indicator and continuous variables are used, with the exception that instead of D_{n-1} (taxiing delay of previous departure), D'_{n-1} was used, which represents the scheduled delay of the last departure aircraft based on runway sequence.

Table 2 presents the results of the regression. Results show that advisory indeed reduces scheduled delay, and the effect is significant. This can be attributed to the use of system delay as the objective in the SARDA scheduler, as discussed in section II.A. However, the effect of advisory in heavy traffic cases is not significant. Another interesting observation is that TMI aircraft have higher scheduled delays; this can be explained from the facts that 1) the aircraft have to wait for the desired take-off time and 2) the scenarios themselves were designed such that TMI times could be achieved, artificially inducing delay.

TABLE 2: RESULTS FROM LINEAR REGRESSION ON SCHEDULED DELAY

Variable	Coefficient Estimate	Std. error	t-value	Pr(>|t|)
Intercept	155.60	38.06	4.09	0.00
A	-85.79	37.84	-2.27	0.02
H	72.90	40.90	1.78	0.07
W	74.52	58.22	1.28	0.20
TMI	384.65	67.73	5.68	0.00
L	0.01	0.01	0.45	0.65
D'_{n-1}	0.58	0.04	13.05	0.00
A×H	8.12	18.66	0.43	0.66
A×W	184.54	27.81	6.63	0.00
A×TMI	56.16	27.64	2.03	0.04
A×D'_{n-1}	-0.11	0.04	-2.79	0.01
A×L	0.02	0.01	1.68	0.09
H×W	-9.89	30.03	-0.33	0.74
H×TMI	183.87	28.97	6.35	0.00
H×L	-0.02	0.01	-1.06	0.29
H×D'_{n-1}	0.02	0.05	0.39	0.70
W×L	-0.03	0.02	-1.38	0.17
W×D'_{n-1}	-0.20	0.06	-3.65	0.00
TMI×L	-0.11	0.02	-5.22	0.00
TMI×D'_{n-1}	-0.27	0.06	-4.44	0.00

G. Controller Workload

While investigating decision support tools for air traffic control, it is necessary to examine the effect of the tool on the controller in terms of usability, workload and acceptance. Data was collected from the participating controllers for this purpose. For the efficiency aspect, real-time workload ratings were taken every five minutes; NASA Task Load Index (TLX) workload ratings [21] were collected at the end of each run, and other post-run questionnaire responses were analyzed to assess how much internal resources (e.g., spare capacity for workload, spare attention) the controllers felt they had during each run. For the satisfaction aspect, the controllers' responses to the post-run and post-study questionnaire regarding their subjective judgment on the helpfulness of the SARDA advisories and ease of use of the user interface were examined.

Detailed results from this analysis can be found in [18]. In summary, the results do not show any increase in workload with the use of advisory. In fact, the NASA TLX workload rating results exhibited clear reductions of workload levels in terms of Temporal Demand (time pressure), Effort (how hard controllers had to work physically and mentally), Physical Demand (e.g., using EFS, communicating on the radio), and Mental Demand (e.g., thinking, deciding, calculating, remembering, looking). In all four ratings, the magnitude of the mean-score reductions from the Baseline runs to the Advisory runs was approximately 2 points, which may have been large enough to be sensed by the controllers.

V. CONCLUSIONS

The 360° simulated tower HITL experiments of tactical gate hold using SARDA showed promising results. Taxiing delay reduction of 60% was observed in heavy traffic scenarios. Estimated fuel reduction of 33% was observed in the heavy scenarios, with emissions showing a similar trend. Reductions in delay and fuel were also observed for the medium traffic level. Even though the amount of gate hold for every aircraft was not limited, less than 5% aircraft in one scenario had a gate hold of more than 15 minutes. Arrival aircraft delay remained unaffected, and the runway throughput with the use of advisory seems the same as the baseline case. TMI compliance was not much different between the advisory and the baseline case. The taxiing delay reductions due to the use of advisory are statistically significant. There is some evidence to suggest that the use of advisory decreases the overall scheduled delay in the system, although further tests are needed to evaluate this.

Although the results are promising, there are limitations in this study. The effect of pushback uncertainty was not studied. Although this has been studied in offline simulations, studying this in a HITL environment would be invaluable in transitioning the concept to any airport. Connected to this is the need for a ramp controller interface, which is currently under development at NASA Ames Research Center. Of course, there is a need for evaluating this concept in a field trial, where many more sources of ramp movement uncertainty are present, including baggage and fuel carts, uncertain pushback paths and others. Lastly, with increasing acceptance of departure metering as a method for reducing surface fuel consumption and emissions, policy debates and changes might be required to allow gate holding without adversely affecting airline performance metrics.

ACKNOWLEDGMENTS

The authors are very grateful to the SARDA simulator team for their hard work and patience: Easter Wang, Dan Pietrasik, Cynthia Freedman, Kenneth Ray, Carla Ingram, Boris Rabin Betty Silva, and David Chin. The authors also wish to thank fellow SARDA researchers for inputs, including Bob Windhorst, Justin Montoya, Victoria Dulchinos and Jon

Holbrook. The authors are also grateful to the controller subject matter experts for advice and participation: Mark, Mitch, Barbara, Fred, Rebecca, and Kari. The authors are also grateful to Mark Hansen and Yi Liu for their participation in the study. Lastly, a big thank you to all the controller and pilot participants.

REFERENCES

[1] A. Nakahara and T.G. Reynolds. "An Approach for Estimating Current and Future Benefits of Airport Surface Congestion Management Techniques," 12th AIAA ATIO and 14th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, 2012. Indianapolis, Indiana, USA: AIAA.

[2] H. Daniel and M. Tim. "Estimating the Achievable Benefits of Airport Surface Metering," 12th AIAA ATIO and 14th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, 2012. Indianapolis, Indiana, USA: AIAA.

[3] C. Brinton, C. Provan, S. Lent, T. Prevost, and S. Passmore. "Collaborative Departure Queue Management," 9th USA/Europe Air Traffic Management Research and Development Seminar (ATM2011), 2011. Berlin, Germany.

[4] Y. Jung, et al. "Performance Evaluation of a Surface Traffic Management Tool for Dallas/Fort Worth International Airport," 9th USA/Europe Air Traffic Management Research and Development Seminar, 2011. Berlin, Germany.

[5] A. Nakahara, T.G. Reynolds, T. White, C. Maccarone, and R. Dunsky. "Analysis of a Surface Congestion Management Technique at New York JFK Airport," 11th AIAA ATIO Conference, 2011. Virginia Beach, VA, USA: AIAA.

[6] I. Simaiakis, et al. "Demonstration of Reduced Airport Congestion Through Pushback Rate Control," 9th USA/Europe Air Traffic Management Research and Development Seminar, 2011. Berlin, Germany.

[7] D. Böhme. "Tactical departure management with the Eurocontrol/DLR DMAN," 6th USA/Europe Air Traffic Management Research and Development Seminar, 2005. Baltimore, Maryland, USA.

[8] R. Deau, J.B. Gotteland, and N. Durand. "Airport surface management and runways scheduling," 8th USA/Europe Air Traffic Management R&D Seminar, 2009. Napa, California, USA.

[9] I. Gerdes and A. Temme. "Taxi routing for aircraft: Creation and Controlling," The Second SESAR Innovation Days, 2012. Braunschweig, Germany.

[10] J. Gulding, et al. "US/Europe comparison of ATM-related operational performance–An initial harmonized assessment by phase of flight," 8th USA/Europe Air Traffic Management Research and Development (ATM2009), 2009. Napa, California, USA.

[11] G. Gupta, W. Malik, and Y. Jung. "An Integrated Collaborative Decision Making and Tactical Advisory Concept for Airport Surface Operations Management," 12th AIAA ATIO and 14th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, 2012. Indianapolis, Indiana, US: AIAA.

[12] W. Malik, G. Gupta, and Y. Jung. "Spot Release Planner: Efficient Solution for Detailed Airport Surface Traffic Optimization," 12th AIAA ATIO and 14th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, 2012. Indianapolis, Indiana, US: AIAA.

[13] G. Gupta, W. Malik, and Y. Jung. "Incorporating Active Runway Crossings in Airport Departure Scheduling," AIAA GNC Conference, 2010. Toronto, Canada: AIAA.

[14] J. Montoya, Z. Wood, and S. Rathinam. "Runway Scheduling Using Generalized Dynamic Programming," AIAA GNC Conference, 2011. Portland, Oregon, USA: AIAA.

[15] Raytheon, CTO-05 Surface Management System, CTOD 24 Final Report, R.A. Team, Editor. 2004, May.

[16] S. Atkins, et al. "Surface management system field trial results," AIAA 4th Aviation Technology, Integration, and Operations Forum, 2004. Chicago, IL, United states: American Institute of Aeronautics and Astronautics Inc.

[17] SAIC, Airspace Traffic Generator: User's Manual Supplement, Revision 2.6. 2006.

[18] M. Hayashi, et al. "Usability Evaluation of Spot and Runway Departure Advisor (SARDA) Concept in Dallas/Fort Worth Airport Tower Simulation," 10th USA/Europe Air Traffic Management Research and Development Seminar (ATM2013), 2013. Chicago, USA., 2013.

[19] T. Nikoleris, G. Gupta, and M. Kistler, "Detailed estimation of fuel consumption and emissions during aircraft taxi operations at Dallas/Fort Worth International Airport," Transportation Research Part D: Transport and Environment, 2011.

[20] S.P. Washington, M.G. Karlaftis, and F.L. Mannering, Statistical and economic methods for transportation data analysis. 2009: Chapman & Hall/CRC.

[21] S.G. Hart and L.E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," Human mental workload, 1988. **1**: p. 139-183.

AUTHOR BIOGRAPHIES

Dr. Gautam Gupta received a B.Tech. in Civil Engineering from the Indian Institute of Technology Bombay, India, and both M.S. and Ph.D. in Transportation Engineering from University of California, Berkeley. He is currently a Research Scientist at UARC-NASA Ames Research Center, researching methods of improving airport efficiency. His research interests include airline operations and planning, airline economics, and large scale optimization problems in transportation.

Dr. Waqar Malik received his B.Tech. degree in mechanical engineering from the Indian Institute of Technology, Kharagpur, India, in 2001 and both M.S. and Ph.D. degrees in mechanical engineering from the Texas A&M University, College Station, Texas, in 2003 and 2007, respectively. He is currently a Research Scientist at University of California, Santa Cruz, where he is involved in developing a surface management system aimed at enabling significant increase in efficiency and throughput of airport surface operations while maintaining safety. He has previously worked on the synthesis of fixed structure controllers, combinatorial motion planning, collision avoidance and path planning for unmanned vehicles.

Dr. Leonard Tobias received his PhD. from the Polytechnic Institute of Brooklyn in 1969. He has been engaged in ATM research at NASA Ames Research Center for over forty years.

Dr. Yoon Jung is an aerospace engineer at NASA Ames Research Center, Moffett Field, California, USA, since 2000, specializing in research on airport surface operation. Prior to joining NASA, he worked at Advanced Rotorcraft Technology and Raytheon. He received both B.S. and M.S. from Seoul National University, Seoul, Korea, and Ph.D. degree in Mechanical Engineering from University of California, Davis. Dr. Jung is a Senior Member of American Institute of Aeronautics and Astronautics.

Ty Hoang received his Bachelor ('91) and Masters ('94) of Science degrees in aeronautical engineering at the California Polytechnic State University at San Luis Obispo. He started his career working on a pseudo-aircraft piloting system then participated in the field-testing of the Traffic Management Advisor (TMA) system at DFW in the 1990's. Next, he assembled the Multi-Center TMA (McTMA) HITL simulator leading to successful field-tests. He now leads the SARDA simulator work.

Dr. Miwa Hayashi is an aerospace engineer at NASA Ames Research Center with background in aerospace human factors. She received a B.Eng. from Tokyo Metropolitan Institute of Technology in 1994, a M.S. in Aerospace Engineering from the University of Texas at Austin in 1997, and a Ph.D. in Humans and Automation from Massachusetts Institute of Technology in 2004. At Ames, she has been helping designing and evaluation of user interfaces for spacecraft cockpit and air traffic controllers' workstations.