Examining Arrival Time Predictability in the NAS using Monte Carlo Methods

Seli Agbolosu-Amison, Ph.D. member AIAA, MITRE Corporation
David C. Millner, MITRE Corporation
Billy Baden Jr., MITRE Corporation
Gareth O. Coville, MITRE Corporation
Stéphane Mondoloni, Ph.D. Associate Fellow AIAA, MITRE Corporation

Introduction and Background

The Next Generation Air Transportation System (NextGen) implementation plan articulates that NextGen will become a better way of doing business because travel will become more predictable [1]. This is important because predictably of flight operational events, in particular flight arrival times, have a significant impact on direct operating cost to the airline industry and passengers and an indirect effect on the economy [2, 3]. Uncertainty in flight arrival times causes carriers to increase their scheduled gate-to-gate times (schedule block times) by adding a buffer (or pad), a practice called schedule padding. Airlines claim that improvements to arrival time predictability could add revenue flights to their schedules without the cost of aircraft acquisitions [4].

There are various mechanisms that produce the arrival time uncertainty. Some of these, such as the variability in airport capacity due to airport configuration and meteorological conditions and the variability associated with positioning in a growing or decaying queue, are considered in fast-time simulations of NAS-wide performance. This paper looks at the additional incorporation of variations that can occur in the times of three processes: pushback-from-gate, taxi-out to the departure queue, and taxi-in to gate (herein called the GTT processes). These variations are incorporated into the *systemwide*Modeler, a fast time discrete-event simulation of the National Airspace System (NAS) that runs nation-wide scenarios to evaluate NAS performance for a full day of air traffic operations. The *systemwide*Modeler was chosen because a recently developed framework for distributing *systemwide*Modeler runs across multiple hosts (the MITRE Elastic Goal-directed (MEG) simulation framework) allows the Monte Carlo simulations needed to build up the distributions required for a predictability analysis to be completed in an acceptable time frame.

Surface movement events such as the GTT processes currently have simplified process models (e.g., average taxi times) in the *systemwide*Modeler. This study replaces the un-modeled portions of the GTT processes (described later) for all flights in individual simulation runs with random variates drawn from distributions estimated from empirical data. A Monte Carlo across runs (the GTT MC) is then used to estimate the variation in flight arrival times. This was done in Agbolosu-Amison [5] by introducing these GTT variates into the simulation outputs on a post-run basis and validated against operational data While the post-run approach allows the estimation of first-order effect of changes of these variations on individual flights, these effects do not propagate to connected flights in this post-run approach. This paper extends the approach by modifying simulation inputs on a pre-run basis to allow modified flights to interact with each other within the simulation in order to investigate flight arrival predictability and to refine the validation work. This is important since for a given day of NAS operation, the variance observed in the real world is influenced by delay propagating between elements in the NAS system. For example, a late arriving aircraft can result in a delayed departure. With certainNextGen operational improvements potentially affecting delay propagation through increased operator flexibility, it is important to ensure that this effect is appropriately captured in the model.

Approach

*Proposed GTT processes under systemwide*Modeler

For a typical scheduled flight from an origin to destination airport, the GTT processes under the version of *systemwide*Modeler utilized for this study have un-modeled effects as follows:

- The pushback-from-gate process does not account for carrier caused delay, and other processes not included in *systemwide*Modeler such as crew availability, baggage delays, etc.;
- The taxi out process does not account for non-departure queue delay (i.e. delay due to interactions during taxi); and
- The taxi-in process does not account for surface arrival delays (e.g., delay for crossing an active runway). This is because *systemwide*Modeler models taxi-in time as a constant for all flights.

In Agbolosu-Amison [5], an approach was taken that follows the same flight (e.g., AALxyz at 13:00 from IAD to ORD, where x, y, z=0-9) across multiple days over a year of ASPM data to develop the standard deviation for the un-modeled part of each GTT process. The averages of the standard deviations of all such flights are the individual variates that form the empirical distributions.

Proposed systemwideModeler architecture

The MEG simulation framework enables an analyst to launch multiple *systemwide*Modeler runs concurrently or asynchronously within a distributed environment. Figure 1 shows the execution of *systemwide*Modeler within MEG. MEG provides for a generalized infrastructure to schedule jobs, monitor, execute, and optimize simulations. MEG can be used to launch multiple replications of a model to a cluster of hardware resources in parallel [6]. The results from completed executions are kept for later analysis and the cycle is repeated until all planned jobs are scheduled and have completed.

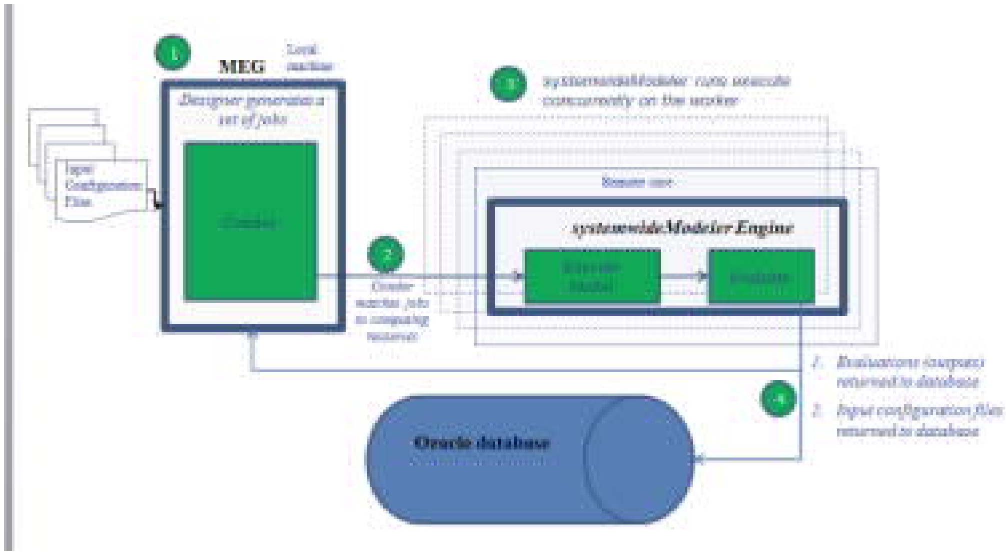

Figure 1: Proposed *systemwideModeler* architecture

Derivation of the empirical distributions

Two sets of GTT empirical distributions will be developed by combining both FY2010 Aviation System Performance Metrics (ASPM) flight data and simulated outputs flight data from *systemwide*Modeler. These distributions will be developed by 1) obtaining modeled GTT process distributions of simulated outputs of NAS processes and 2) determining the un-modeled GTT processes distributions empirically by taking the difference between random correlated samples generated from both respective simulated and observed empirical distributions. The first set will be conditioned on airport. Agbolosu-Amison [5] demonstrated an improved system-wide predictability performance by conditioning on airports. A second set will be developed conditioned on airport and season (the monthly quartiles of the fiscal year: fall, winter, spring and summer). This will be done to investigate if further improvement is possible. Figure 2 shows examples of the two sets of distributions developed for the taxi-in process for Atlanta.

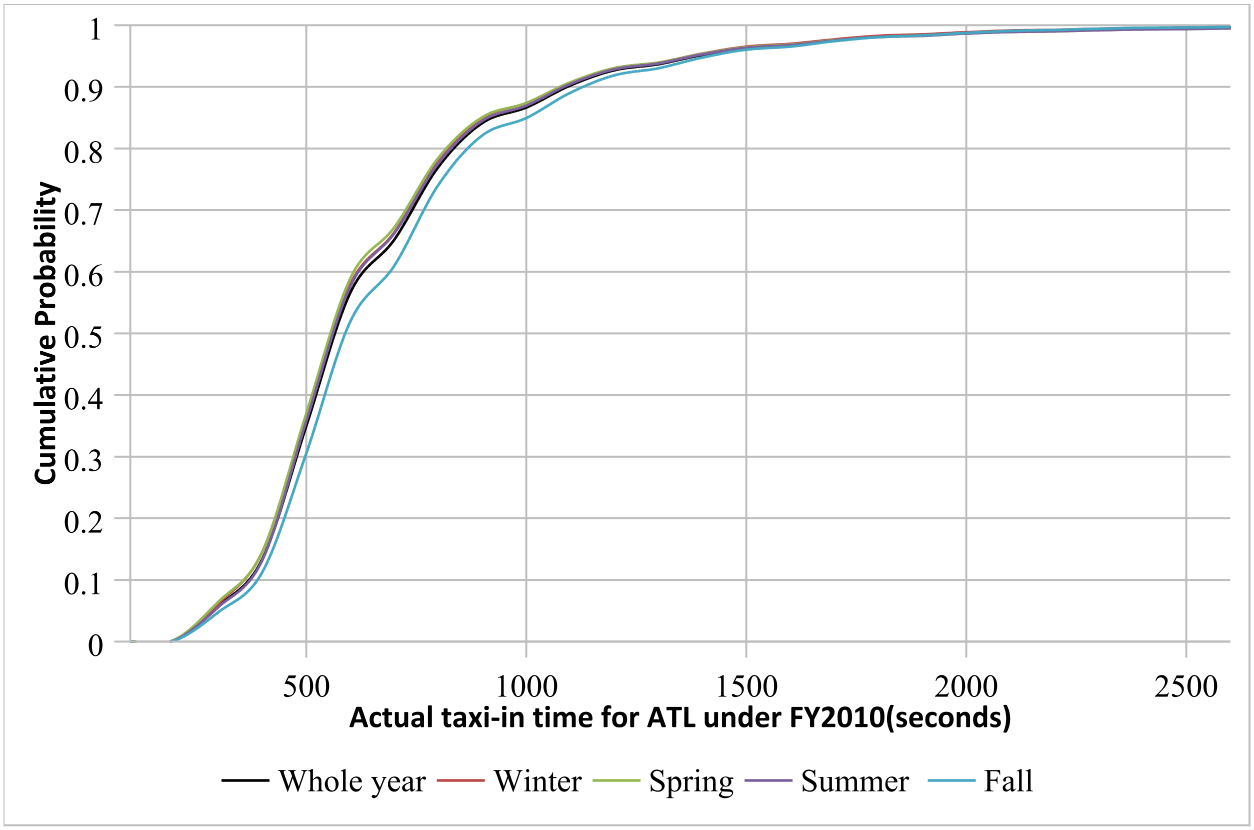

Figure 2: Taxi-in time empirical distributions for ATL under FY2010

Design of simulation runs

A GTT MC is a set of 3,600 runs of the*systemwide*Modeler. Each run combines one of a selected set of 36 scenario days from FY2010 with one of 100 different sets of random variates for the GTT process. The scenario days were selected using a process that balances airport operations, airport weather and en route center operations across the NAS [7]. Each set of random variates uses a different random number seed. Table 1 below shows the four different GTT MCs used.

Table 1: Gate and Taxi Time Monte Carlo Cases

GTT MC Cases	Within model or Post-processing	
Empirical GTT Distributions (annual or seasonal)	Pre-run + By-Year	Post-run + By-Year
	Pre-run + By-Season	Post-run + By-Season

In addition to the four GTT MCs, a separate set of Base Runs is also usedfor the validation analysis. These consist of one run per scenario day using the baseline GTT process times.

Processing of simulation results

Predictability Analysis

For this analysis, we will investigate predictability through analysis of the variation in the following delays:

- Arrival Delay: measures how much after the nominal at-gate time that a flight arrives.
- Schedule Delay: measures how much after the scheduled at-gate time that a flight arrives.

These variations can occur as a result of effects that are predictable and those that are not. For example looking at the above delays for the NAS as a whole, variations occur for a variety of reasons. Examples of predictable effects include:

- Variations in the flight time as a result of aircraft type, stage length, or average seasonal wind effect.
- Variation in the taxi-out time as a result of the average taxi-out time experienced at a departure airport, by a carrier (due to location of their gates) and at a specific time of day (due to congestion).
- Variation in the taxi-in time for similar reasons to taxi-out (i.e.,arrival airport, gate location, average runway crossing delay).

Effects that are not predictable, at the time of schedule formulation, include:

- Variation in the flight time as a result of winds on a specific day, or route choice given conditions including avoidance of congestion or convective weather.
- Variations in the pushback and taxi-out times as a result of unscheduled maintenance, timing of arrival into a growing or decaying departure queue, or specific surface interactions at congested times.
- Variations in the taxi-in time as a result of surface congestion, specific runway crossing circumstances, or gate unavailability.

We separate effects into those that are predictable at the time of schedule formulation from those that are not at that time because our goal is to determine the impact of predictability changes on the additional required schedule padding to achieve a target on-time performance. Effects that can be accounted for in the scheduling will be reflected in the mean, whereas targeting on-time performance above the 50^{th} percentile requires additional padding related to the magnitude of the uncertainty.

We will apply linear mixed effects statistical models [8, 9] to separate the contributions of predictable variation from those that remain unpredictable once the predictable variation has been accounted for. The output of this approach provides functions enabling us to investigate the changes in the statistics of on-time performance as NextGen operational improvements are incorporated into the simulation model. This provides an estimate of the additional schedule padding required to achieve a target on-time performance.

Validation Analysis

Our validation analysis will seek to determine whether the pre-run improves upon the post-run approach, or provides results that are comparable in quality. This is achieved through a comparison of various validity measures. Three different measures of validity that measure the difference between scheduled-out and gate-in time are defined relative to ASPM data:

- Base Validity: Base runs compared to ASPM.
- Post-GTT Validity: post-run GTT MC compared to ASPM.
- Pre-GTT Validity: pre-run GTT MC compared to ASPM.

Since not all modeled airports are contained within the ASPM airports, the data will be segregated into three groups corresponding to both, one or no ASPM airports being either origin or destination airport of the flight. Only results for the first two groups can be compared to ASPM data.

As for the predictability analysis, we will apply mixed effects statistical models to estimate the mean validity in each of the three cases described above. In an improved model, one would expect the mean validity to improve. One also would expect consistency in the magnitude of effects (random or fixed) and improved or constant unexplained variation.

Previous results [5] have provided good NAS-wide results for on-time performance estimation using a post-run approach. The validation analysis seeks to ensure that the results of the pre-run approach are consistent or improved over the previous approach.

Proposed results and conclusions

Research is still ongoing and the expected final results will include metrics from post-run and pre-run of *systemwide*Modeler under the GTT MC simulations. These will indicate whether inclusion of un-modeled effects through pre-processing and conditioning distributions on season and airport provides valid results. The inclusion of effects through pre-processing is not expected to improve upon the validation data in [5], but provides a model capable of propagating effects of changes in the GTT processes enabled by select surface management improvements. Output metrics will illustrate the validity of this modified *systemwide*Modeler for modeling and measuring system-wide predictability.

References

1. Federal Aviation Administration, *NextGen Implementation Plan*, March 2011.

2. Ball, M., Barnhart, C., Dresner, M., Hansen, M., Neels, K., Odoni, A. et al., "*Total Delay Impact Study: A Comprehensive Assessment of the Costs and Impacts of Flights Delay in the United States*", Final Report, Washington, D.C..: Federal Aviation Administration, October 2010.

3. Wandeler, Y. D. and Marsh, D. "*Planning for Delay: Influence of Flight Scheduling on Airline Punctuality*". EUROCONTROL Trends in Air Traffic Volume 7. 2008

4. ATSP Focus Group and CNS/ATM Focus Team. "*Airline Metric Concepts for Evaluating Air Traffic Service Performance*". February, 1999

5. Agbolosu-Amison, S. J. and Mondoloni, S., "*Modeling System-wide Predictability and Associated Air Carrier Benefits*", AIAA-2011-6810, AIAA Aviation Technology, Integration and Operations (ATIO) Conference, including the AIAA Balloon Systems Conference and 19th AIAA Lighter-Than, Virginia Beach, VA, September, 2011

6. Wickham, B., Agbolosu-Amison, S.J., Baden Jr., W., Litwin, L. "*Efficient Search of a NAS-WIDE Analysis Space*". 2012 Integrated Communications Navigation and Surveillance (ICNS) Conference, April 2012, Virginia.

7. Baden Jr., W., Bodah, D. J., Williams, A. G., Kuzminski, P. C. "*SystemwideModeler: A Fast-Time Simulation of the NAS*". 2011 Integrated Communications Navigation and Surveillance (ICNS) Conference, May 2011, Virginia.

8. Bates, D., "*Tutorial: Fitting and Evaluating Mixed Models Using lme4*" http://user2010.org/tutorials/Bates.html, The R User Conference, July 2010.

9. Goldstein, H., *Multilevel Statistical Models, 4th Edition*, Wiley, 2011.