# Impact of Select Uncertainty Factors and Implications for Experimental Design

Gareth O. Coville<sup>1</sup>, Billy Baden, Jr.<sup>2</sup> and Rishi Khanna<sup>3</sup> *The MITRE Corporation, McLean, Virginia 22102, United States* 

Performance estimates produced by a National Airspace System (NAS)-wide simulation models vary due to the complexity and amount of variability that occurs within the NAS. One area of modeling variability which current NAS-wide simulation models attempt to compensate for is the variation in delay occurring across days. This is typically accomplished through the use of a carefully-selected set of days seeking to be "representative" of the NAS performance across a given year. These days are referred to as design days. Current practices model each design day once, with averaging across all design days to yield annual estimates of performance. The concern with this process is that each design day represents one specific instance of what could have happened in the NAS and does not consider the many small daily variations that could have a potentially significant impact. Also, creating design days is an interactive and time-consuming process, so simply creating additional design days to improve the confidence of the model results is not always economically feasible. This paper determines the impact that intra-day perturbations due to four factors within a NAS-wide model (air carrier delay, runway configuration changes, sector workload limits and program rate forecasts associated with ground delay programs) have on NASwide simulation results. This paper also determines the combinations of design days and iterations per design day required to achieve convergence of NAS-wide estimates for a given confidence level when the four factors within the model are perturbed. The conclusions of this paper are that averaging across design days provides a high level of confidence in the results up to a point but for even higher levels of confidence it becomes important to include iterations in the experimental design. The four factors added an additional 37% NAS-wide delay to the model results. We expected the four factors to increase delay in model results as some of these factors were not previously modeled and are new forms of delay. The factor that contributed the most to the variability of NAS-wide delay was the program rate forecasts associated with ground delay programs.

### I. Motivation

As Air Navigation Service Providers (ANSP) across the globe invest in and deploy operational upgrades in alignment with the ICAO Global ATM Operational Concept [1], performance justification is required for these investments [2]. Pre-deployment, these are frequently justified at a national or regional scale through modeling and simulation (M&S) activities to estimate the performance impact of operational improvements. Often, these improvements can deliver value through relatively small changes in performance thereby necessitating M&S capabilities capable of differentiating the effect of the proposed improvement from the effect of modeling variability.

One area of modeling variability which current M&S practices attempt to compensate for is the variation in performance occurring across days [3-4]. This is typically accomplished through the use of a carefully-selected set of days seeking to be "representative" of the National Airspace System (NAS) performance across a given year.

<sup>&</sup>lt;sup>1</sup> Project Team Manager, Performance & Economic Modeling & Analysis, 7525 Colshire Drive McLean, VA 22102 /Mail Stop N590.

<sup>&</sup>lt;sup>2</sup> Lead Multi-Discipline Systems Engineer, Performance & Economic Modeling & Analysis, 7525 Colshire Drive McLean, VA 22102 /Mail Stop N590.

<sup>&</sup>lt;sup>3</sup> Intern, Performance & Economic Modeling & Analysis, 7525 Colshire Drive McLean, VA 22102 /Mail Stop N590.

These selected days are referred to as design days. Simulations are then performed on the design days and statistics are obtained to characterize the performance under a baseline and a treatment case. Current practices model each day in the set of design days once, with averaging across all design days to yield annual estimates of performance.

The concern with this process is that:

- 1. Each design day represents one specific instance of what could have happened in the NAS on that day and does not consider the many small perturbations which impact the NAS daily (e.g. an aircraft taking off five minutes later due to a slow boarding process or an aircraft's trajectory being adjusted because runway configurations were changed ten minutes earlier than expected). Certain design days, such as days when portions of the system are operating at or near capacity, are expected to be more sensitive to these perturbations than other design days.
- 2. Creating design days is an interactive and time-consuming process, so simply creating additional design days to improve the confidence of the model results is not always economically feasible. However, it may be possible to improve the confidence in model results by perturbing available design days multiple times instead of creating new design days.
- 3. Some input factors are not available and not modeled, so increasing the number of design days will not capture their effects on the output.

When providing perturbed input to individual design days, one must decide how many iterations of each design day to perform in order to obtain suitably narrow confidence intervals for any chosen confidence level.

# **II.** Objectives

The objectives of this report are to:

- 1. Identify the impact that intra-day perturbations and fidelity of select input data have on NAS-wide simulation estimates.
- 2. Determine the combinations of design days and iterations (i.e., perturbations of each design day) required to achieve convergence of NAS-wide estimates for a given confidence level.

# III. Approach

*systemwide*Modeler is a fast-time, discrete-event simulation tool developed by the MITRE Corporation to simulate air traffic and its interactions with various elements of the NAS[5]. A typical *systemwide*Modeler simulation consists of tens of thousands of flights progressing along four-dimensional trajectories, responding to constraints imposed by capacity-limited resources like airports and en route sectors. Because *systemwide*Modeler is a deterministic model, uncertainty must be represented through changes to model inputs over a series of simulation runs. With run-times in the tens of minutes, performing a sufficient number of simulation runs to assess the impact of uncertainty on NAS performance has previously been impractical. Advances in multicore architectures have enabled parallel processing of simulation runs on a large enough scale to facilitate Monte Carlo analysis by varying model inputs[6]. Model inputs now no longer need to be treated as constants but instead can be randomly selected from statistical distributions, providing a greater understanding of their impact on model results and the relationship the input variables have with one another.

## Perturbations contributing to model sensitivity

This paper examines the delay impact of four factors of uncertainty upon NAS performance: (1) air carrier delays, (2) ground delay program (GDP) arrival rate forecasts, (3) timing of airport configuration changes and (4) sector workload limits. These four factors are known to affect system performance, but processing constraints previously limited the extent to which they could be studied. Now that *systemwide*Modeler can facilitate a much larger number of runs on a parallel processing architecture, the ability exists to more accurately measure the effect that these factors have on simulated performance by assigning a distribution to each factor and conducting a Monte Carlo analysis.

Air carrier delays and delays due to GDP arrival rate forecasts are new inputs of delay not previously modeled in *systemwide*Modeler. They were expected to increase mean delays in the model results. The timing of airport configuration changes and sector workload limits were already represented in *systemwide*Modeler, so adding variability to these factors was expected to increase the variance of system performance but was not expected to affect mean delays in the model results.

To represent the impact of these factors on NAS performance, input parameters to systemwideModeler were varied with each simulation run. A distribution for each parameter or set of parameters was created based on

historical data sources, such as reported carrier-induced delays, or through an understanding of errors being introduced, such as for the timing of configuration changes as explained below. For each design day being simulated, perturbations are created by randomly selecting from the distributions for the parameters in order to provide perturbed input to the model.

# Air carrier delays

Fig. 1 illustrates air carrier delay distributions for six of the Core 30 airports which were obtained using air carrier delay data from the Bureau of Transportation Statistics (BTS) databases for the seven largest US carriers. The air carrier delay data collected by BTS represents delay "due to circumstances within the airline's control (e.g. maintenance or crew problems, aircraft cleaning, baggage loading, fueling, etc.)."<sup>4</sup> The figure describes the air carrier delay for only those flights that were subject to a delay. To approximate these distributions, an empirical NAS-wide carrier delay distribution is applied. The NAS-wide carrier delay distribution was created by combining the distributions from the six individual airports into a generic distribution. By assuming that air carrier delay are imposed prior to the flight pushing back, we can draw from the distribution to determine how much air carrier delay to impose upon flights prior to the execution of the simulation. It is recognized that carrier-induced delay may occur post-pushback in the real world, but for the purposes of simulating this delay we imposed it pre-pushback in the model. The air carrier delay to be imposed on each flight is the result of two independent draws, one to determine if a flight is subject to a carrier delay at all (10 percent chance), and a second to determine the delay from the distribution.[8]

# **GDP** arrival rate forecasts

A GDP is implemented at an airport due to inclement weather, an aircraft incident, closed runways or other factors that may impact arriving traffic. A GDP reduces the actual rate at which aircraft arrive at the airport. In anticipation of the reduced airport capacity, a program rate is issued when a GDP is first scheduled. The program rate is equivalent to the forecast of the Airport Acceptance Rate (AAR) used by Air Traffic Control (ATC) to manage arrival flows to the destination airport.

A GDP is issued when the ATC System Command Center (ATCSCC) broadcasts an advisory, which defines the GDP program parameters, along with a dynamic flight list informing system users (e.g. airline dispatchers and ramp controllers) and the departure Air Traffic Control Tower (ATCT) of the specific Expected Departure Clearance Time (EDCT) for each flight being delayed by the GDP. The program parameters in the advisory include the planned GDP start time, planned GDP end time, and the program rate associated with each hour during the program. The departure airport ATCT is expected to manage the traffic flow so that each flight with an EDCT departs the airport within a few minutes of that EDCT time.

Because the program rate issued by the advisory is an independent forecast involving professional judgment, there is a possibility that the rate may not correspond to the actual AAR in effect during the GDP. The ratio of the program rate to the AAR is called the GDP arrival rate ratio. It is an indication of how accurately the program rate forecasts the AAR. A GDP arrival rate ratio of one indicates that the program rate was the same as the AAR. A GDP arrival rate ratio noe indicates that the program rate underestimated or overestimated the AAR, respectively.

To model the GDP arrival rate ratio, we collected five years of data from the FAA's National Traffic Management Logs (NTML) and Aviation System Performance Metrics (ASPM) databases. From this data we determined that eight airports were impacted by GDP's for more than 2.5 percent of the hours over the five years. These airports are Newark Liberty International (EWR), LaGuardia (LGA), San Francisco International (SFO), Chicago O'Hare International (ORD), John F. Kennedy International (JFK), Philadelphia International (PHL), General Edward Lawrence Logan International (BOS) and Hartsfield-Jackson Atlanta International Airport (ATL). The GDP patterns for each airport are unique, so to model the GDP arrival rate ratio, we created airport-specific distributions using the following steps:

1. Determine if a GDP occurs at an airport based on number of hours of Visual Meteorological Conditions (VMC) for the day. We determined that the likelihood of a GDP occurring on a specific day is highly correlated to daily VMC hours, with correlation values ranging from -0.85 (ATL) to -0.94 (BOS).

<sup>&</sup>lt;sup>4</sup> Research and Innovative Technology Administration, U.S. Department of Transportation, "Understanding the Reporting of Causes of Flight Delays and Cancellations",

http://www.bts.gov/help/aviation/html/understanding.html#q3, August 16, 2012

- 2. Calculate the length of time that the GDP will be in effect based on the number of VMC hours for the day. We found the daily VMC hours was strongly correlated to the length of time that a GDP is implemented, with correlation factors ranging between -0.71 (SFO) to -0.93 (PHL).
- 3. Calculate the start time of the GDP based on the time of day. We found that the start time of a GDP is highly correlated to the time of day, with GDP's typically occurring in the afternoon.
- 4. Calculate the length of time between when the initial advisory was sent by the ATCSCC and when the GDP was implemented. We found that the initial advisory was often issued at the same time as the GDP was being implemented, but could be issued up to six hours in advance of the GDP start time.
- 5. Calculate the GDP arrival rate ratio. By comparing the program rate from the NTML data with the AAR published in ASPM, we were able to determine the GDP arrival rate ratio. We anticipated that the program rate would become more accurate the nearer the forecast was made to the actual GDP implementation time. However, this was not always the case. For example, at EWR the average GDP arrival rate ratio remains close to 0.9 even when the forecast time is very small (or even equal to zero), as shown in Fig. 2 below. We believe that EWR under-estimates the program rate for tactical reasons including having the capacity to deal with unscheduled pop-up flights. Similar observations were made at the other airports. However, the forecast under-estimation patterns were not as distinct as at EWR.



Figure 2 – Mean GDP Arrival Rate Ratio at EWR for 2007 - 2011

The output from these steps was an event log listing the airport, time of initial forecast, GDP start time, GDP end time, and GDP arrival rate ratio. The event log is created at the start of each model run and used as an input for *systemwide*Modeler.

Within *systemwide*Modeler, a Merging and Spacing (MAS) resource is used to deliver flights to an arrival airport according to some arrival rate. A separate resource within *systemwide*Modeler compares anticipated airport demand and capacity profiles to determine a schedule of arrival rates to be used by MAS. The schedule of arrival rates used by MAS is modified by the GDP arrival rate forecast in order to model the overestimate or underestimate of the GDP program rates.

# **Timing of runway configuration changes**

Ceiling, visibility, and approach information provided by ASPM is used to determine which runway configuration is in effect each quarter hour. Due to the discrete nature of the data, the actual time that a runway configuration could have changed is equally likely to have happened anytime within the quarter hour. The timing of runway configuration changes is important because it can have a significant impact on the airport's capacity. Airport configurations are not explicitly represented in *systemwide*Modeler. Instead, representative configurations describe the airport capacity in visual, marginal (MMC) and instrument meteorological conditions (IMC). Therefore, a uniform distribution over the quarter hour leading up to the configuration change was used to adjust the time of configuration changes as scheduled in *systemwide*Modeler input.

## Sector workload limits

The number of flights that a sector can handle is dependent on the workload associated with each flight. Each flight entering a sector has some set of tasks that must be performed by the sector air traffic control team. Each of those tasks has a task time. Those task times are summed to build a profile of the work that the air traffic controller must carry out over time. When the projected amount of work that would have to be performed by the controller exceeds the capacity to do work, flights requesting to enter the sector must be delayed in an upstream sector until the controller can accommodate the workload created by that flight. In *systemwide*Modeler, the parameter governing the controller's workload limit was drawn from a distribution of historical workload data that identified what the workload levels were when the sector control needed to switch from two to three controllers. The distribution is provided in Fig. 3 and the average sector workload limit was calculated to be 795 seconds, which was the sector workload limit used for all sectors in the baseline runs.

# **Design of Experiments**

Using *systemwide*Modeler, six cases were tested – a baseline case where no factors were perturbed, one case for each of the four factors previously described, and a "vary all" case where all four factors were perturbed simultaneously. A single model run was conducted for each of the 36 design days in the baseline case. For the remaining five cases where factors were perturbed, 100 iterations of each of the 36 design days were run for each case. This resulted in 18,036 runs of *systemwide*Modeler. For each run, the NAS-wide delay was measured and summarized in the "Results" below.

# **IV.** Results

#### Model sensitivity to individual perturbations

Fig. 4 below summarizes the results from all of the simulation runs. The results in Fig. 4 have been sorted by design day and delay. From this figure it is possible to compare the variability of model results across design days when no factors are perturbed (i.e. the baseline case) as well as the variability to delay results due to the 100 perturbations of each factor across each design day. There is a large range of delays between design days, even before any perturbations are included. For example the baseline delay for Design Day 23 (3.97 minutes/flight) is roughly three times the amount of delay for Design Day 34 (1.32 minutes/flight). We can see that the timing of runway configuration changes has only a small impact on model results, and that the GDP arrival rate forecasts add the largest amount of variability to model results of the four factors. The air carrier delay and sector capacity factors have less impact on model results than the GDP arrival rate forecasts, but these two factors add delay that was not previously simulated.



Figure 4 – Impact of Individual Runs on systemwideModeler Results

Fig. 5 below is a boxplot showing the range in results when each factor is perturbed independently. The impact from design day variability was removed by calculating the change in delay as a percentage of baseline delay. The whiskers of the boxplot represent the minimum and maximum percentage of delay, the top and bottom of the box represent the 25<sup>th</sup> percentile and 75<sup>th</sup> percentile percentages of delay and the line through the box represents the median.



Figure 5 – Impact of Factors on systemwideModeler Delay

Air carrier delays and GDP arrival rate forecasts are causes of delay not modeled in the baseline and so add to the output NAS-wide delay – on average 19 percent and 5 percent of additional delay per model run, respectively. The configuration change and sector capacity (workload limits) were modeled in the baseline, but without any distribution applied to them. While the timing of configuration changes had little effect on NAS-wide delay, when the sector workload limits were perturbed, using the distribution outlined in Fig. 3 above, the resulting delay was always higher than the delays calculated in the baseline case. On average the sector workload limit perturbations added an additional 12 percent of delay to the model results. Sectors can be thought of as a series of queues that aircraft need to pass through when traveling to their desitnation airport. Therefore a distribution applied to the sector workload limit will mean that some sectors will be assigned a workload limit lower than average. These sectors act as bottlenecks and the backup they cause has ripple effects throughout the NAS. We confirmed this result by rerunning the analysis with an adjusted sector workload limit distribution that had the same mean as the original distribution, but a standard deviation reduced by 50 percent. Again delay was consistently added to the results, but the increase was 2 percent of delay - significantly less than the 12 percent of delay obtained from the original distribution. Sectors with low sector workload limits have a large impact on delay because rerouting functionality was not used in the model runs. Rerouting is available to real-world operations, so the extent of this delay is likely a modeling artifact of systemwideModeler.

The vary all case highlights the sensitivity of model results when all four factors are varied together and independently. For each factor, the perturbations applied to each iteration are identical in the factor-specific case to the perturbations applied in the vary all case. For instance, if flight 1 is given 45 seconds of pushback delay in iteration 1 of the air carrier delay runs, it is also given 45 seconds of pushback delay in iteration 1 of the vary all runs. (In contrast, flight 1 is given no pushback delay in iteration 1 of the sector workload limit runs because air carrier delay is not the factor being varied.)

In the vary all case, NAS-wide delay averages 37 percent more than in the baseline case. This is one percent greater than the sum of average percentage of delay from the carrier induced (19 percent), GDP arrival rate forecasts (5 percent), sector workload limit (12 percent) and runway configuration (0 percent) delays. The standard deviation of delay added by the vary all case is 0.2 percent larger than the standard deviation of the delay added by the four factors independently. It is not surprising that the delays added by these four factors are largely independent of one another. The four factors generally affect different parts of the system being modeled. Air carrier delay affects delay at flight pushback. Airport configuration change times affect airport capacities. Sector workload limits affect congestion in en route sectors. Only GDP arrival rate forecasts directly affect elements of the system directly impacted by other factors. The GDP arrival rate forecasts affect the delivery rate of arrivals to the affected destination airport. If the timing of configuration changes is varied at the same time as the GDP arrival rate forecast, both could be affecting the airport capacity. Because the impact of the timing of configuration schedule changes is small, the interactive effects are negligible. Likewise, if the GDP arrival rate forecast results in a less densely populated arrival stream, more delays may need to be absorbed prior to pushback from the origin airport, which could interact directly with the pushback delay imposed on flights to represent air carrier delay.

# Model sensitivity to all perturbations

Now that a greater number of *systemwide*Modeler simulations can be performed simultaneously by using parallel computing capabilities, conducting large numbers of iterations in conjunction with multiple design days has become a feasible option for introducing uncertainty. Therefore we want to understand how the sensitivity of NAS-wide delay results can be better managed using combinations of design days and iterations (i.e., number of perturbation runs per design day).

We did this by determining the range of possible average NAS-wide delay results when using a specific number of design days and iterations. We used the 3,600 model runs calculated in the vary all case as our sample from which to draw.

First, we determined the range of possible average NAS-wide delay results when varying the number of design days and keeping the number of iterations constant at one. We did this by randomly selecting different groups of a set number of design days, and choosing one iteration for each design day in the group. Next, the average of the NAS-wide delays across each iteration in the group was calculated. The value for the average NAS-wide delay varies based on the design days and iterations selected. The set number of design days was varied from one to 36 and the process was repeated 1,500 times per set of design days, giving us (36 \* 1,500 =) 54,000 average NAS-wide delay values to work with. We then calculated the 5<sup>th</sup> percentile and 95<sup>th</sup> percentile average NAS-wide delays from the 1,500 results for each set of design days and presented the results in Fig. 6 below. When reviewing these results

it is important to remember that this study was limited to a total of 36 design days and had to assume that the "true mean" is the average NAS-wide delay taken from the 3,600 vary all model runs.

As expected, the ranges of possible average NAS-wide delay results converge as the number of design days that make up the average increases. For example, when only using one model run of one design day from the vary all case, the range of possible results between the 5<sup>th</sup> and 95<sup>th</sup> percentile is anywhere from 1.51 to 5.14 minutes/flight of NAS-wide delay, but taking the average delay of one iteration across 36 design days reduces this range to between 3.00 to 3.18 minutes/flight. The range of possible values when using the baseline case is superimposed onto Fig. 6 to highlight the delays added to the *systemwide*Modeler results due to air carrier delay, GDP arrival rate forecasts and the variation of sector workload limits. The dotted lines in the Fig. 6 represent the "true mean" for the baseline case (2.29 minutes/flight) and the vary all case (3.08 minutes/flight).



Figure 6 – 5<sup>th</sup> and 95<sup>th</sup> Percentile Range of Average NAS-wide Delay Results when Design Days in the Sample are Increased (Iterations = 1)

We are also interested in determining if simulating more iterations of a smaller set of design days can still achieve the same range of possible average NAS-wide delay results, since it is now cheaper to complete multiple iterations than to create additional design days. Fig. 7 below shows how the sensitivity of the model results decrease as the number of iterations is increased. We used the same process to generate the results presented in Fig. 7 as we did for Fig. 6, with the only difference being that we varied the number of iterations (between 1, 10, 50 and 100) as well as the design days. Similar to Fig. 6, this figure shows the 5<sup>th</sup> and 95<sup>th</sup> percentile range of all possible average NAS-wide delay values resulting from a given number of design days and iterations. From this figure we can see that increasing the number of iterations has relatively little impact on the range of model results when the number of design days is small.



Figure 7 – 5<sup>th</sup> and 95<sup>th</sup> Percentile Range of Average NAS-wide Delay Results when Design Days and Iterations in the Sample are Increased

However, once the number of design days increases sufficiently, the number of iterations becomes relatively more important, as shown in Fig. 8 below. Fig. 8 represents exactly the same data as shown in Fig. 7 for 30 through 36 design days except that the vertical scale has been changed to improve readability. This figure shows that, after 30 design days, the number of iterations has an increasingly greater impact on reducing the range of possible results, until the "true mean" is reached at 36 design days and 100 iterations. This suggests that the range of possible results would be better reduced by increasing the number of iterations than by increasing the number of design days once the set of design days consists of at least 30 days. For example, the range between the 5<sup>th</sup> and 95<sup>th</sup> percentile average NAS-wide delay when using one iteration of 36 design days (0.183 minutes/flight) is *worse* than using 100 iterations of 33 design days (0.169 minutes/flight) or 10 iterations of 34 design days (0.171 minutes/flight).



Figure 8 – 5<sup>th</sup> and 95<sup>th</sup> Percentile Range of Average NAS-wide Delay Results for 30 to 36 Design Days when Iterations in the Sample are Increased

After evaluating every combination of design days and iterations, we were able to create the contour chart in Fig. 9 below. Fig. 9 shows the combination of design days and iterations needed to reach a given coefficient of variation in model results. The coefficient of variation is calculated by dividing the standard deviation by the mean of all possible results for a specified number of design days and iterations. A smaller coefficient of variation indicates a smaller range of possible values relative to the mean and a greater level of confidence that the calculated average NAS-wide delay will be relatively near to the mean.



Figure 9 – Coefficient of Variation for Design Day and Iteration Combinations

### V. Conclusions

This paper set out to determine the impact that intra-day perturbations and fidelity of select input data have on NAS-wide simulation results, and the combinations of design days and iterations required to achieve convergence of NAS-wide results within a specified level of confidence.

This paper identifies the uncertainty associated with adding four NAS-wide factors that increase delay in a simulation - air carrier delay, GDP arrival rate forecasts, runway configuration change times and sector workload limits. These factors were found to add delay to a design day of 19 percent, 5 percent, 0 percent and 12 percent, respectively. When all factors were varied simultaneously, it was found that delay increased by 37 percent and that the primary effects that the four factors have on the system being simulated are nearly independent of one another. We found that most of the variation in NAS-wide delays resulted from the GDP arrival rate forecast.

Previously we had kept capacities constant across all sectors, but found that NAS-wide delays increased when we added variation to sector workload limits. Sectors can be thought of as a series of queues that aircraft need to pass through when traveling to their desitnation airport. Therefore a distribution applied to the sector workload limit will mean that some sectors will be assigned a workload limit lower than average. These sectors act as bottlenecks and the backup they cause has ripple effects throughout the NAS. Sectors with low sector workload limits have a large impact on delay because functionality was not used in the model runs to allow flights to reroute through adjacent sectors. Rerouting is available to real-world operations, so the extent of this delay is likely a modeling artifact of *systemwide*Modeler. Additional investigation is warranted to confirm that the delays due to the sector capacity perturbations are removed when the rerouting functionality is switched on in *systemwide*Modeler.

Using the uncertainty of these four factors, we were able to show how the range of possible average NAS-wide delay results converges as more design days are used. We were also able to show that, after 30 design days, the number of iterations has an increasingly greater impact on reducing the range of possible results, until the "true mean" is reached at 36 design days and 100 iterations. For this study, "true mean" was defined as the average NAS-wide delay taken from the 3,600 vary all model runs. This suggests that the range of possible results would be better reduced by increasing the number of iterations than by increasing the number of design days once the set of design

days consists of at least 30 days. For example, the range between the 5<sup>th</sup> and 95<sup>th</sup> percentile average NAS-wide delay when using 1 iteration of 36 design days (0.183 minutes/flight) is *worse* than using 100 iterations of 33 design days (0.169 minutes/flight) or 10 iterations of 34 design days (0.171 minutes/flight). This is important because now that parallel processing computer architectures have reduced the processing time of each simulation run it has become much cheaper to run multiple iterations than to create new design days.

Finally, we created a chart (Fig. 9) to show the relationship between design days and iterations needed to achieve a given level of confidence in NAS-wide simulation results.

#### Acknowledgments

The authors would like to thank Dr. Seli Agbolosu-Amison, Wayne Cooper, Stanley Mejia, David Millner, Dr. Stephane Mondoloni, Lorrie Smith, and Brian Wickham. All are employed at The MITRE Corporation and this paper would not have been possible without the contributions each made in their specific area of expertise.

## Disclaimer

The contents of this document reflect the views of the authors and The MITRE Corporation and do not necessarily reflect the views of the U.S. Federal Aviation Administration or the U.S. Department of Transportation. Neither the FAA nor the DOT makes any warranty or guarantee, expressed or implied, concerning the content or accuracy of these views.

#### References

[1] ICAO, Global Air Traffic Management Operational Concept, ICAO Doc 9854, First Edition, 2005.

[2] ICAO, Manual on Global Performance of the Air Navigation System, ICAO Doc 9883, First Edition, 2008.

[3] Foster, G., "Demand Generation for System-Wide Simulation", 2<sup>nd</sup> Annual Workshop on Innovations in NAS-Wide Simulation in Support of NextGen, Center for Air Transportation Systems Research, George Mason University, January 2010

[4] Gulding, J., "ATO Future Schedule Generation", 2<sup>nd</sup> Annual Workshop on Innovations in NAS-Wide Simulation in Support of NextGen, Center for Air Transportation Systems Research, George Mason University, January 2010

[5] Baden, B., Bodoh, D., Williams, A., G. and Kuzminski, P. C., "systemwideModeler: A Fast-time Simulation of the NAS", *I-CNS Conference*, Herndon, VA, May 2011.

[6] Wickham, B., Agbolosu-Amison, S., Baden, B., Litwin, L., and Smith, L., "Efficient Searching of a NAS-wide Analysis Space", *I-CNS Conference*, Herndon, VA, April 2012.

[7] Agbolosu-Amison, S. and Mondoloni, S., "Modeling System-wide Predictability and Associated Air Carrier Benefits", *ATIO Conference*, Virginia Beach, VA, September 2011.

©2012 The MITRE Corporation. All rights reserved.