*Proceedings of the 1988 Winter Simulation Conference*
M. Abrams, P. Haigh, and J. Comfort (eds.)

# Validation of a simulation model of the National Airspace System

Melvyn Cheslow
The MITRE Corporation
McLean, Virginia 22102

**ABSTRACT**

A validation program has been developed to check the accuracy and validity of a simulation model being developed for the Federal Aviation Administration. The program is composed of four components: conceptual model validitation, input data validation, computerized model correctness checks, and operational validation. As part of the operational validation, model output is compared with three aviation delay databases which vary in their definitions and measurement of delay. Therefore, a major part of the validation program has been to determine adjustments that must be made to model output in order to make valid comparisons.

## 1. INTRODUCTION

This paper describes the validation program for a simulation model of the National Airspace System (NAS). The model was developed for the Federal Aviation Administration (FAA) to describe, in an aggregate manner, the effects of the interactions of the many components of this airspace system without including so much detail that the data collection and model run times would be prohibitive. Therefore, the challenge has been to represent the important aspects of the system without a detailed representation of the activities at individual airports or the actions of individual air traffic controllers.

The validation of this aggregate model, hence, could not focus on the extent to which the model accurately simulates each aircraft activity, but instead had to test whether the model is correct in a more aggregate sense. The validation program has been further limited by the lack of any specific data collection or experimental efforts. Instead, only readily available data have been used. Especially important are several aviation databases that provide delay estimates used extensively by aviation analysts. However, these databases are not necessarily consistent in their delay estimates; they are collected by different organizations and measure different phenomena. Therefore, a major part of this validation program has been to define precisely what types of delays are measured, where the delays occur, and what adjustments have to be made to the model outputs in order to make valid comparisons.

## 2. THE SIMULATION MODEL

Under the direction of the FAA, the MITRE Corporation has developed a prototype performance model of the National Airspace System. This model is being developed in two phases. The first phase provides a deterministic simulation of the air traffic using 58 airports and the local airspace around some of the airports. The second phase expands the model to include major portions of the airspace in the United States and adds stochastic elements. This paper discusses only the validation of the Phase 1 model.

The model is a high-level simulation of all of the flights on a given day using the 58 modeled airports. Approximately 40,000 scheduled and non-scheduled flights are simulated. Actual aircraft schedules are used, including the extensive peaking of operations that is currently in practice. The simulation model is described in a paper presented at this conference by Weiss and Lacher (1988).

The major inputs of the simulation model are as follows:

1. The scheduled departures and arrivals, and the combining of all flight legs by the same aircraft into "itineraries".
2. The unscheduled operations by general aviation (including business aircraft), and military aircraft.
3. The service rates at airports for departures and arrivals. There is a tradeoff in the model between serving departures and arrivals, and the service rates depend on the relative demands.
4. The service times at departure and arrival "fixes". These are major locations, where the air traffic control system controls aircraft merging and spacing after airport departure or before airport arrival.
5. The weather pattern across the nation during the day. The airport service rates, as well as the number of operations at an airport, depend on the weather. Commonly, the service rates are classified as either Visual Flight Rules (VFR) or

Instrument Flight Rules (IFR).

## 3. THE VALIDATION PROGRAM

Validation of the model is part of an iterative process. The process starts with data collection and model development; validation then occurs. If necessary, changes to the model or data are suggested, and improvements to the model or data are made. This loop continues until the model appears accurate enough to be used for specific analyses. The analyses also provide new information about model accuracy that either indicate further improvements to the model, or suggest that additional validation tests be performed.

The validation program for these models followed the literature, especially the book by Law and Kelton (1982) and the paper by Sargent (1987) at last year's Winter Simulation Conference. In particular, the nomenclature for the four components of the validation program are derived from Sargent's paper. These are:
-- conceptual model validation
-- input data validation
-- computerized model verification
-- operational validation

Conceptual model validation involved testing both the logical structure and reasonableness of the model. The model had to be consistent with physical and engineering facts such as locations, distances and speeds and with air traffic control procedures. A basic requirement was that simplifications in the model should produce small or unimportant effects on outputs.

Input data validation involved checking data for reasonableness and consistency. The data should be obtained from reliable sources. However, in practice, some data are incomplete or out of date. For example, the Official Airline Guide (OAG) provides information about all scheduled flights, but does not indicate the number of canceled flights. Therefore, the OAG somewhat over-estimates aircraft activity on bad weather (IFR) days. Also, the FAA publishes the number of daily operations of unscheduled aircraft for every airport, but does not report on hourly activity. The only source for these latter data are from airport records.

Data items were checked for internal consistency. For example, capacity estimates for various arrival and departure fixes were initially obtained from different local ATC staff members who made differing assumptions. Similarly, data for different, but related, parameters in the model were checked for reasonableness. Fix capacities were compared with runway capacities to ensure they were the same order of magnitude; the number of scheduled and non-scheduled operations at each airport were compared to see that the major, congested airports had small proportions of non-scheduled operations during peak periods.

Computerized model verification involved checking the system design and code for errors. The system design was examined by outside experts, and traces of the code were run to ensure that each portion of the code worked properly.

The fourth component, operational validation, involved testing for valid model output. This included checks for reasonableness, as well as simple statistical comparisons of delay and flow estimates with other databases. Operational validation has been a major effort for this modeling program because of the existence of several databases that provide delay estimates extensively used in analyzing aviation services. Also, as discussed in Section 8, below, operational validation is an on-going activity while the model is used for various analyses.

Simple reasonableness checks were performed such as the following:
- boundedness -- ensure that no negative delays are produced; ensure that airport utilization is always less than one
- degeneracy -- see that eliminating aircraft does not increase delay
- extreme input behavior -- ensure that increasing airport and fix capacities eventually eliminates delays

As part of the reasonableness checking, graphical presentations of several of the model outputs were examined visually for anomalies. This examination is an important aspect of the validation since it can be easily performed by managers and operations experts.

Simple statistical checks included two types of tests, calibration and sensitivity analysis. Calibration involves checking model outputs, and then modifying the model so that the outputs "match" some prescribed values. The required level of the match is discussed below. Calibration can be a mindless exercise in which "fudge-factors" are used to ensure that the match occurs. The more appropriate approach, however, is to examine the mismatches for patterns, and then hypothesize a cause. The cause may be an error in either input data, model algorithm, or model code. It may, alternatively, be an error in the data set used for the calibration. This portion of the validation effort requires the talents of a detective.

A hierarchy of requirements was developed for the calibration effort. At the most aggregate level, total system-wide delays should be matched. Next, in order, came individual airport delays, delay components such as arrival and departure delays, and distributions of the length of delays.

Sensitivity analysis involves analyzing the differences in model output before and after an input parameter is changed. Examples are the closing of a runway, increasing capacity at a single airport, and smoothing schedule peaking.

Precise statistical tests were not performed in this validation program because of the admonitions of Law and Kelton (1982) that autocorrelation in both the "real world" and model output make "classical statistical tests...not directly applicable."

Instead, an accuracy criterion has been used that suggests that the model should be as accurate as the databases it is compared against, and accurate enough to analyze the particular issues it is used for. The model should be able to discriminate between two alternatives, or to compare a new situation with a base case.

To ascertain how accurate a model must be, the accuracy of the delay databases that it is compared against must be determined; uncertainty of the order of 20% -40% does not appear unreasonable for the national delay databases. It is likely that the accuracy requirements for analyses of issues will vary depending on the specific analyses. A model can never be considered absolutely valid. The goal for this validation effort is to ensure that the model is "good enough".

## 4. VALIDATION DATABASES

Three aviation delay databases have been used to validate the model outputs. These are:

- -- FAA National Airspace Performance Reporting System (NAPRS)
- -- FAA Standard Air Carrier Delay Reporting System (SDRS)
- -- Department of Transportation Airline Service Quality Performance File (ASQP)

NAPRS delays are reported daily by ATC facilities to FAA headquarters by cause and by time of day for all IFR traffic. The reports are modified and corrected by headquarters personnel to attribute causes outside of the reporting facility, e.g., flow control restrictions imposed by the local center or by the Central Flow Control Facility. Delays are reported for three location types, departure (gate and taxi-way combined), en route, and arrival (last center or terminal airspace). Only delays greater than or equal to 15 minutes at each location are reported; these delays are called "reportable delays". NAPRS can miss cumulative delays of more than 15 minutes for a single flight if each facility that the flight traverses sees less than a 15 minute delay. The baseline for reported delays is judgmental; zero delay is not a well-defined condition.

SDRS delays are based on data reported monthly by three airlines to the FAA for each of their flights. Average delays per aircraft at each airport are calculated by FAA staff and used to estimate delays for all scheduled operations. Delays are categorized by five locations: departure gate, departure ramp, departure taxi-out, en route and taxi-in. They are calculated relative to baselines established by each airline. En route baselines are the airlines' standard, no-delay, times for each flight segment, by aircraft type. Taxi-out and taxi-in baselines are the 10th percentile shortest times for each airport and aircraft type. With the latter definition, 90% of all operations will have these ground delays. Gate delays begin when an aircraft is ready for push-back. Therefore, whenever an aircraft is not available to leave the gate at its scheduled time due to its late arrival,the resultant delay is excluded from the SDRS.

The ASQP delays are reported by 13 airlines to the Department of Transportation (DOT) monthly for every flight. Departure and arrival delays are measured relative to scheduled times for departing and arriving at the gate. Therefore, ASQP departure delays may include delays due to previous late arrivals. But, they omit any delays between leaving the gate and wheels off. ASQP arrival delays, on the other hand, include this last category of delays, even though they occur at the departure airport. The ASQP file probably has the most extensive and most accurate information about airline delays, but its classification scheme is not consistent with many other definitions. The ASQP arrival delays, however, do correspond with the simulation model's "effective delay" --the delay that is perceived by the passenger-- and therefore are very useful in the calibration program. (These modeled delays are defined further in the next section.) The ASQP reports also include information about cancellations and diversions that are due to reasons other than mechanical problems.

## 5. VALIDATION MEASURES

Several measures have been used in the quantitative comparisons of model output with the delay databases. These include:

- -- number of daily operations
- -- total daily delay in minutes at each queue
- -- average delay per aircraft at each queue
- -- hourly delays
- -- distribution of the length of delay
- -- number of delays greater than 15 minutes
- -- effective delay: the passenger's perceived arrival delay

The effective delay is the difference

between a flight's actual time of arrival and its scheduled time of arrival. These delays provide a measure of the magnitude of the "ripple effect", the delays of an aircraft that accumulate due to an inability to make up delays occurring earlier in the day.

Arrival and departure delays at each airport have been examined both separately and combined. With this information, it could be determined, for example, whether capacities appeared to be in error or whether the arrival/departure trade-off algorithm needed adjusting. Of course more significant problems could also cause poor fits, such as too simple a representation of the runways, or omitting important components of the airports. Delays caused by airport- and fix-capacity limitations also have been compared separately and combined, to obtain insights about the relative importance of fix-related delays.

Some commonly produced output measures from simulations have been less useful in the validation program than the above measures. Queue length and throughput have been used to ensure the code is performing correctly, but have not been used in the operational validation because of the lack of any field data on these measures.

## 6. DATA ADJUSTMENT

In order to use the delay data to validate the simulation model, the categories and the definitions of delay had to be equalized as much as possible. For comparisons with NAPRS data, this meant using the number of delays greater than 15 minutes. Since NAPRS only includes delays by IFR aircraft, model output for a VFR-day scenario exceeded the reported delays, especially for airports with a high proportion of general aviation, such as Teterboro or Santa Ana.

For the SDRS data comparisons, all departure delays in the SDRS were combined; however, these categories include some delays due to ground congestion not represented in the model. SDRS arrival delays were comparable to the modeled arrival delays, but included some non-modeled en route delays. (NAPRS data indicated that the delays caused in the en route airspace are relatively small, but NAPRS does not include all en route delays, only those due to weather and to aircraft placed in holding patterns.) SDRS taxi-in delays were omitted, since these kinds of delay were not modeled. Special model output was created with a post-processor to report the average aircraft delays by the airlines reporting in SDRS, and these were used in the validation.

The ASQP arrival delays were compared with the modeled effective delays. ASQP departure delays, which occurred only at departure gates, were not used, because they could not be made comparable to any component of the modeled delays.

## 7. VALIDATION RESULTS

The calibration of model output has been accomplished for a VFR day and is currently being performed for several weather scenarios, each of which represents different mixes of capacities. In addition, several sensitivity analyses have been performed involving changes to capacities, traffic levels, and Air Traffic Control (ATC) operating procedures such as departure restrictions.

Some significant model improvements came about from the early quantitative model calibration. By examining the magnitude of arrival and departure delays during periods when one or the other demand was heavy, it was determined that the arrival/departure tradeoff rule should be improved and made more realistic. Additionally, comparisons of modeled airport delays with SDRS data indicated that the capacities (inverse of service rates) of some airports should be modified. Also, detailed examination of the number of operations and average delays over arrival and departure fixes determined that some of the assignments of aircraft to departure fixes were incorrect; and sensitivity analysis of the importance of departure restrictions at runways due to fix capacity limitations indicated that the levels of the restrictions at some airports were wrong.

## 8. ON-GOING PROGRAM

The validation program is expected to continue beyond calibration as long as the simulation model is used to analyze new problems and alternative solutions. Each analysis will require a specific validation to ensure that, at the very least, alternative solutions can be correctly ordered with respect to their relative improvements.

## REFERENCES

Law, A.M. and W.D. Kelton (1982). *Simulation Modeling Analysis*, McGraw-Hill Book Company.

Sargent, R.G. (1987). An Overview Of Verification and Validation of Simulation Models. In: *Proceedings of the 1987 Winter Simulation Conference* (A. Thesen, H. Grant, and W.D. Kelton, eds.) Atlanta, Georgia, 33-39.

Weiss, W. and E. Lacher (1988). Simulating the National Airspace System. In: *Proceedings of the 1988 Winter Simulation Conference*. San Diego, California

## AUTHOR'S BIOGRAPHY

Melvyn Cheslow is a Member of the Technical Staff in the Air Transportation Systems Engineering Division of the MITRE Corporation, where he is currently involved in developing and using a simulation model of the National Airspace System. He received a BS in Physics at the California Institute of Technology and an MS in Computer Science from Johns Hopkins University. He started his professional career as a solid state physicist, modeling electro-optical devices such as lasers and solar cells, and radiation effects on materials.

He has been analysing issues in the transportation field for over 20 years. He has developed demand forecasting models for transportation services and for automobile ownership and useage, and was part of a team that simulated the impacts of alternative solutions to transportation problems in the Northeast Corridor of the U,S., including high speed rail, short take-off and landing aircraft, and highway information systems. He recently was involved in modeling air traffic controller tasks and the impacts of new ATC computers.

Melvyn D. Cheslow
The MITRE Corporation
7525 Colshire Drive, W335
McLean Virginia 22102
(703) 883-6764