Limits to Growth: Results from the Detailed Policy Assessment Tool

Frederick Wieland

Center for Advanced Aviation System Development

The MITRE Corporation, M/S W281

1820 Dolley Madison Blvd., McLean, VA 22102−3481, USA

ABSTRACT

Congestion−induced delay is one of the recognized metrics for assessing performance in the National Airspace System (NAS). Delay can arise from many causes, including factors specific to airline operations as well as factors that can be traced to NAS operations. In this paper we introduce a recently developed tool, the Detailed Policy Assessment Tool (DPAT), and use it to explore delay and the predictability of delay using some simple scenarios for NAS growth.

INTRODUCTION

Congestion−induced delay is a significant element in NAS operations and is a major contributor to the cost of airline operations. It has been estimated, for example, that flight delays cost commercial airlines operating in the United States an excess of $3 billion per year in operating costs alone [1]. This estimate excludes the indirect costs to the ultimate users—business people and the travelling public—who must plan for or deal with these resulting delays. With growth in air travel expected to continue well into the next century, it is important to assess the impact of such growth on NAS operations in general, and the delay metric in particular, so that plans to mitigate the delay can be made and ultimately realized.

The MITRE Corporation’s Center for Advanced Aviation System Development (CAASD), a Federally Funded Research and Development Center sponsored by the Federal Aviation Administration (FAA), has recently built a fast−time simulation model capable of estimating terminal−area and en−route delays for hundreds of airports worldwide. This model, called the Detailed Policy Assessment Tool (DPAT), is the focus of this paper and is used to generate simple capacity/delay results presented later. Although DPAT is similar to the NASPAC (National Airspace System Performance Analysis Capability) model, also developed at CAASD [2], there are substantial differences, especially with regard to enroute modeling, worldwide capability, and total execution time.

We shall begin with a discussion of the delay metric in order to define what it is we are measuring, followed by a description of the conceptual model used by DPAT to describe NAS operations, concluding with some simple representative results regarding future NAS delay and throughput. It is beyond the scope of this paper to present specific options or recommendations for mitigating the future delays predicted herein; we merely outline some characteristics of NAS growth and ways in which DPAT can measure the delay impact of such growth.

BACKGROUND

"Delay" is a fundamental quantity used to characterize the performance of all transportation systems—surface as well as airborne. It has historically been used to quantify air traffic performance, and is currently included in the Airline Service Quality Performance (ASQP) database collected by the US Department of Transportation.

Delay is not the only measure of NAS system performance; other measures, as outlined in [3], are equally important and are often overlooked. Nevertheless, it remains a critical performance quantity whose measurement and prediction is essential for assessing the current NAS and options for future configurations.

One of the problems with the delay metric is its definition. To an airline user (a passenger), delay is simply the difference between the scheduled arrival time (SAT) and the actual arrival time (AAT). This definition suffers from some severe shortcomings if adopted as a NAS performance metric. For starters, the SAT is merely an artifact of airline scheduling constraints; it can sometimes be changed to either absorb delay or exacerbate it. Secondly, it masks certain delay−inducing events during the lifetime of a flight, such as takeoff delays, sector handoff delays, or delays due to flow control operations such as ground delay programs or miles−in−trail restrictions. Nevertheless, measurement of this delay is important and we shall define it as the "effective delay."

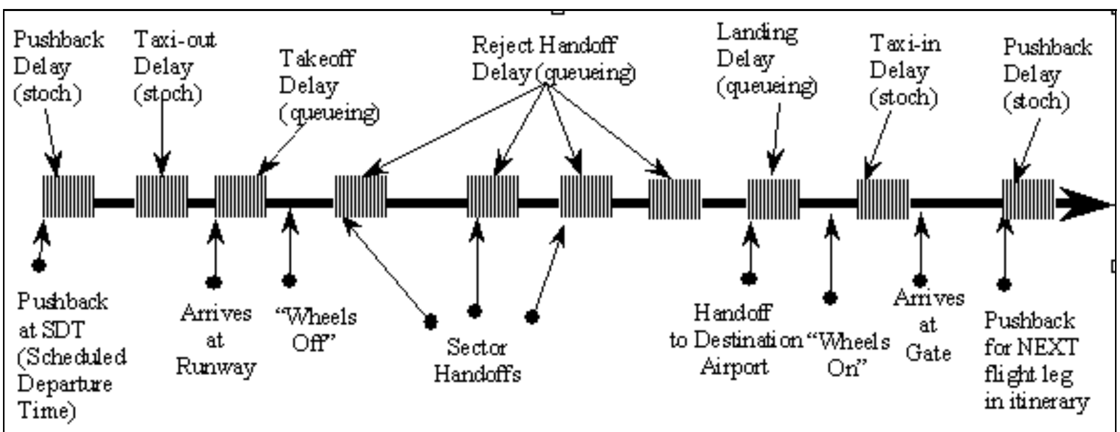

Figure 1. The evolution of a flight as simulated by DPAT.

A better definition of delay is the difference between the actual operating time and the aircraft’s "optimal" operating time. Its "optimal" operating time is the time the aircraft would take, from pushback at the originating airport to arrival at the destination airport, if there were no other aircraft ahead of it and all flow−control mechanisms are absent. We shall call this type of delay—the difference between actual and optimal operating time—the "technical delay." There are various methods to determine optimal operating time, some of which rely on data mining while others rely upon the modeling of aircraft performance; DPAT relies upon a combination of the two.

Note that these two types of delay—effective delay and technical delay—can often be very different. For example, an airline might realize that a mid−afternoon flight to a busy airport is habitually late, assigning it a late SAT—yielding a small or zero effective delay. Its habitually late arrival time might in turn be caused by congestion enroute which forces controllers to slow the plane, vector it, or stack it, yielding a nonzero

technical delay. On the other hand, a late−arriving flight that had been delayed earlier in its itinerary might encounter an uncongested destination airport (with zero technical delays), but arri
ve at the gate very late (nonzero effective delay).

These two types of delay, technical and effective, can be measured separately for arrivals and departures. An airport can have technical arrival and departure delays, while a flight might have an effective departure delay followed by a series of technical delays at the departing airport, enroute, and at the arrival airport, culminating in an effective arrival delay at its destination. Of these two delay measures, technical delay captures the performance of the NAS, while effective delay captures the expectations of passengers. DPAT is configured to measure both types of delay.

Another characteristic of NAS delay is its propagation from one region to another. If there is a congestion−induced bottleneck at one region of the NAS—the Midwest, for example—then larger−than−normal technical delays (arrival, departure, and enroute) can cause flights within the affected region to experience large effective delays. Because arriving flights are often linked to later departing flights, late−arriving flights frequently cause subsequent flights to depart late and, in turn, land late at their respective destinations. In this manner delays at one region can affect distant airports.

THE DPAT MODEL

All simulations are based upon an underlying conceptual model that is itself an abstraction of a real−world system. DPAT is no exception. Because the primary quantities we are concerned with are the measurement and prediction of technical and effective delay, DPAT conceptualizes the NAS as a network of capacitated resources, where the resources are airports, enroute airspace sectors, and enroute fixes. The NAS is represented as a series of queues, each with an appropriately assigned service time, for which airplanes must compete in order to complete a flight.

The queueing abstraction of the NAS captures fairly well the congestion−induced delays experienced by flights. To capture other types of delay, such as those caused by airline operations, DPAT allows the analyst to specify a cumulative distribution function (CDF) describing the stochastic (that is, randomly chosen) delay for various events along the evolution of a flight. For example, by specifying a CDF for pushback events, an analyst can model delay that might be caused by local airline operations. These delays will add to the effective (but not necessarily the technical) delay encountered by the flight.

The evolution of a flight at DPAT’s level of resolution is shown in figure 1. DPAT simulates a flight from gate pushback to gate arrival, abstracting many elements in between. Lacking a detailed airport model, DPAT handles neither gate assignments nor the movement of airplanes along taxiways. Instead, it represents taxi times (both arrival and departing) as a randomly chosen sample from a user−specified CDF that can be tailored to airports and/or airlines. Departure queue and arrival queue times are explicitly computed. Enroute sector handoffs are also modeled, and when the number of aircraft resident in a sector exceeds its user−specified capacity, handoff rejections occur. If the user has specified a ground delay program or miles−in−trail restrictions, the appropriate delays are computed and added to the flight time.

EVALUATION OF THE MODEL

The conceptual model captures the congestion−induced delay while providing a mechanism for adding other sources of delay to a flight. The model’s main strength lies in its ability to propagate delays from one region of the system to another, and that it provides a system−wide view of the impact of delay. Its system−wide view can be very useful for assessing the global impact of proposed local changes that affect capacity or demand in some fundamental way.

Because DPAT is a queueing model, flights exist in one queue or another, rather than at a precise point in the airspace. Its chief disadvantage is that metrics which require precise four−dimensional information—such as proximity metrics or conflict detection/resolution metrics—cannot be estimated by DPAT.

IMPLEMENTATION OF DPAT

DPAT is implemented as a fast−time parallel discrete−event simulation, using advanced parallel processing techniques that have been developed within the last decade [4,5]. The standard hardware configuration uses a Sun SPARCstation 20 with four 120 MHz CPU's and 128 Mbytes of main memory. Using this hardware, DPAT computes the effective and technical delays for the top 500 NAS airports over one simulated day (about 50,000 total flights) in under three minutes. This fast processing time enables dozens of runs to be completed in a short time, enabling analyses that were heretofore impractical.

In addition to the CONUS configuration, DPAT can be configured to simulate all flights worldwide and compute the delay statistics for the top airports in North America, Asia, and Europe. With such a configuration, DPAT simulates about 80,000 worldwide flights in about five minutes.

DATA SOURCES

DPAT is a user−configurable model requiring data on the key quantities of interest—airports, enroute airspace, and enroute fixes. The two most important input items are the capacity and demand data. In this context, "capacity" is a rather fuzzy quantity that is defined differently for airports than for enroute airspace sectors. For airports, "capacity" is simply the number of aircraft that can arrive or depart per hour—what queueing theorists would call the service rate. For enroute airspace sectors, "capacity" refers to the number of flights that can be simultaneously resident in a sector before controllers begin rejecting handoffs. To a queueing theorist, a sector is modeled as a G/G/N/N queue with blocking, where N (the number of servers as well as the maximum number that can be in the sector simultaneously) is identified as the "capacity" of the sector.

In this study, baseline airport capacities for the top twenty NAS airports were extracted from FAA estimates used at the Air Traffic Control System Command Center (ATCSCC). Baseline sector capacities for the 700+ NAS enroute sectors were determined from the FAA's Monitor Alert Threshold (MAT) program, which contains nominal quantities at which sectors are considered overloaded.

A second key datum is the pattern and quantity of air traffic throughout the system, which we call the *demand*. For this study, we used baseline demand extracted from a Thursday in mid−April, 1996, from the Official Airline Guide (OAG). Unscheduled General Aviation (GA) traffic was added to this baseline demand, according to a nonhomogeneous Poisson process whose parameters (rates, airport transfer probabilities) were derived from analyzing 220 days of recorded data at the top 500 NAS airports. Future projections of demand were derived using the Future Demand Generator, a CAASD model that generates additional flights according to user−provided airport and/or system growth parameters.

AIRPORT CAPACITY ALGORITHM

One of the most important algorithms in DPAT is its handling of airport capacities. Because DPAT does not explicitly simulate individual runways, the capacities are expressed as the total arrival and departure rate that a particular airport can handle. Each airport is modeled as a pair of servers each with a single queue, one serving arrivals and one for departures. The service rate for arrivals is dependent upon that for departures, and vice−versa, as follows.

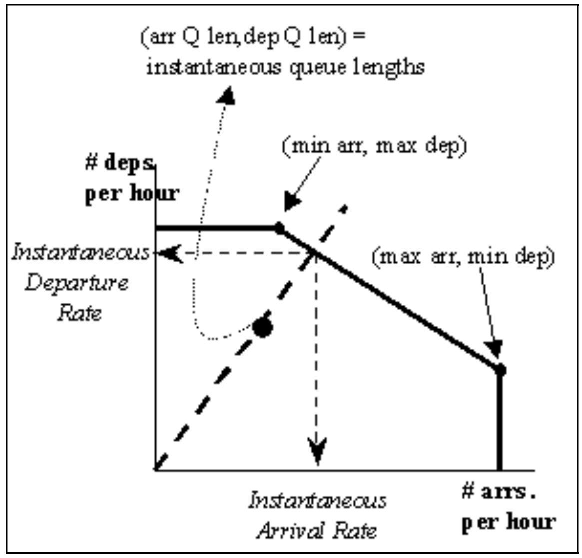

Figure 2. The Airport Capacity Model

Airport controllers regularly configure runway operations to favor one operation or the other, especially during arrival pushes (where arrivals greatly exceed departures) or during departure pushes (vice−versa). To capture the airport dynamics, DPAT uses a capacity model derived from [6], and based upon more than a decade of research by MITRE staff and others. In this model, arrival and departure capacity are linearly dependent upon each other.

Briefly stated, the capacity model is defined by two points in arrival/departure space, one point at (maximum arrival rate, minimum departure rate) and another point at (minimum arrival rate, maximum departure rate). These two points define a polygon in the arrival/departure graph as shown in figure 2. At a given instant in the simulation, when an aircraft requests either an arrival or a departure operation, a third point is formed at (arrival queue length, departure queue length). A line is drawn from the origin through this third point until it intersects the polygon; the arrival and departure rates at the point of intersection are used for the requested operation. This model captures the dynamics of airport capacity tradeoffs while avoiding explicit modeling of runway operations. Its advantage is that the data can be derived either from empirical observations or from fundamental considerations regarding runway configuration, winds, aircraft mix, and so on.

NAS CAPACITY/DELAY STUDIES

As defined, DPAT can be used to perform a number of "what−if" studies of NAS behavior. The set of experiments we chose are neither exhaustive nor are they necessarily realistic, however, they represent some simple DPAT capabilities and provide a quick look at the future NAS. It is not the purpose of this study to propose a particular NAS configuration nor to solve future capacity or demand problems, rather, we shall focus on insights that DPAT can provide given various assumptions. If an analyst wishes to study a particular NAS configuration, s/he can assess its impact on capacities and then can use DPAT to measure the airport−specific and/or system−wide delay effects.

All of the experiments described below use technical delay as the underlying metric, because, as noted earlier, technical delay captures the performance of the NAS without regard to airline scheduling constraints. Technical delay is a congestion−induced queueing delay so these experiments are measuring congestion in the NAS caused by air traffic demand competing for limited resources.

For the first experiment, we held total demand constant at April, 1996 levels, and increased the airport capacities uniformly throughout the system. The baseline capacities (as mentioned earlier) are from the ATCSCC. The technical delay for the top fifty−eight airports (arrival + departure delay) is shown in figure 3.

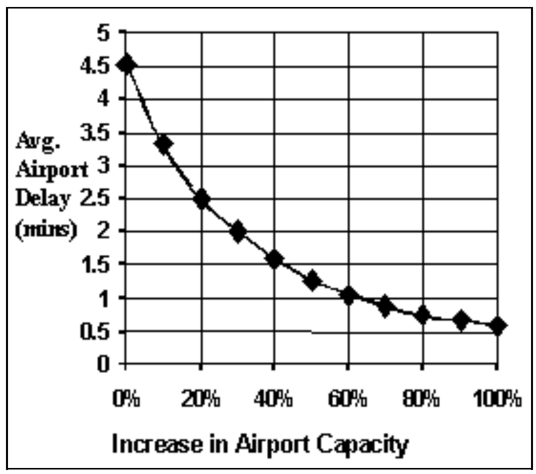

Figure3. Airport Capacity/Delay Analysis

Figure 3 reveals a monotonically decreasing function, possibly with an asymptote (at zero?). Its general shape is consistent with DPAT’s queueing system view of the NAS The result suggests that a uniform doubling of the airport capacities (which will not happen in practice) will reduce delay by about a factor of nine, or almost an order of magnitude.

We can compute a similar curve for the enroute airspace sectors, as shown in figure 4. In this figure we have degraded sector capacity uniformly, holding demand constant at April 1996 levels. The total sector handoff delay is measured, which represents the number of minutes aircraft waited before being accepted into the next sector.

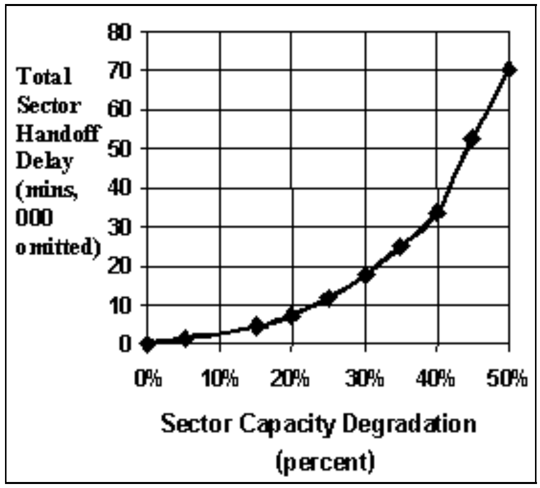

Figure 4. Sector Capacity/Delay Analysis

As expected, the curve is monotonically increasing, whose general shape belies DPAT’s underlying queueing model. The "knee" of the curve can be placed between 20−30%, implying that beyond a 25% reduction in sector capacity the system deteriorates rapidly. Although a uniform decrease in sector capacities would not actually happen in practice, DPAT could be configured to change capacities of individual sectors or regions of airspace, or to compute delay for a different airspace sectorization altogether.

A second interesting set of experiments involves increasing the total system demand (number of flights) while keeping system capacity constant. We can artificially introduce extra flights into the baseline scenario by using the Future Demand Generator (FDG). The FDG requires the growth rate as an input parameter, and it increases traffic consistent with the pattern of traffic within the baseline system demand. In computing figure 5 we assumed that capacities remain constant at the current baseline, and that demand is the only variable that grows. The baseline demand was for October 1995 (as opposed to the April 1996 used for the previous two experiments). The average system−wide delay for the top 58 airports is plotted.

The results suggest that a 25% increase in demand cause the delays to increase by about a factor of two. The shape of the curve, once again, is consistent with the queueing nature of the model.

VALIDATION

As interesting as these results might be, the question remains, how valid are they? In validation, we are concerned with determining whether the conceptual model is an accurate representation of the physical system [7]. While there are many ways to validate a model, we shall be concerned with two primary methods: theoretical and empirical validation. In theoretical validation, we ask whether the conceptual model makes sense, in theory, given what we know about the physical system. In empirical validation, we compare actual results between the simulated and the real system.

First let us consider whether or not the conceptual model is theoretically valid. Because our primary goal is to compute congestion−induced delay and throughput in the NAS, modeling it as a network of capacitated queues seems appropriate. Technical delay, after all, concerns aircraft waiting for scarce resources (runways and sector controllers), which is appropriately modeled as a network of queues. We have already dealt with the shortcomings of this approach vis−à−vis questions that require positional accuracy, but are there any shortcomings with regard to delay calculations?

Indeed there are. The conceptual model currently does not capture weather−related delay effects, such as rerouting around severe weather or outright cancellation of flights. Secondly, and more importantly, the conceptual model is silent about an airline's *response* to NAS disruptions. If sixty−minute delays are predicted at a major airport, then quite probably many airlines will cancel and/or reschedule flights, thus somewhat alleviating the original delay problem.

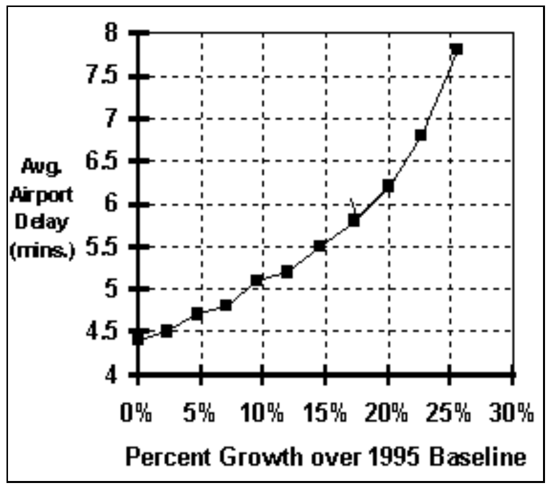

Figure 5. Future Growth Analysis

This airline/NAS interaction effect—where a problem within the NAS causes airlines to adjust, removing the original problem only to reveal new ones—is an effect not captured at all by the conceptual model. Therefore the results presented earlier can only be interpreted as *worst−case* results: in practice, if delays get as large as predicted, the system (airlines, controllers) will adjust so that such delays are never actually realized. It should be noted parenthetically that CAASD is currently working on approximate models of airline behavior, so that DPAT's conceptual model will eventually capture this dynamic effect.

Empirical validation is made difficult by two factors. The first is that the airline/NAS interaction effect remains unmodeled but is realized in practice, so we are comparing "worst−case" predictions against "less−than−worst−case" reality. The second factor is that some of the delays measured by DPAT are hard to measure empirically, and when data are available they are usually noisy and subject to some interpretation.

Despite these difficulties, the validation of DPAT has been addressed in two studies. A comparison of five NAS traffic models was conducted by CAASD [8] with the previously validated NASPAC model used as the baseline. DPAT results were found to track well with those of NASPAC. A second study is currently being conducted to evaluate DPAT validity empirically, by comparing predicted and realized throughputs (rather than delay) at a set of major NAS airports during busy periods of a particular (good weather) day. While the study is still in progress and the results are highly tentative, the preliminary findings suggest that, with an appropriate choice of airport capacities on a clear day (no weather effects), the predicted and realized throughputs are within 20% of each other for about 75% of the observations. Such a result is quite respectable for a model containing some of the shortcomings mentioned earlier.

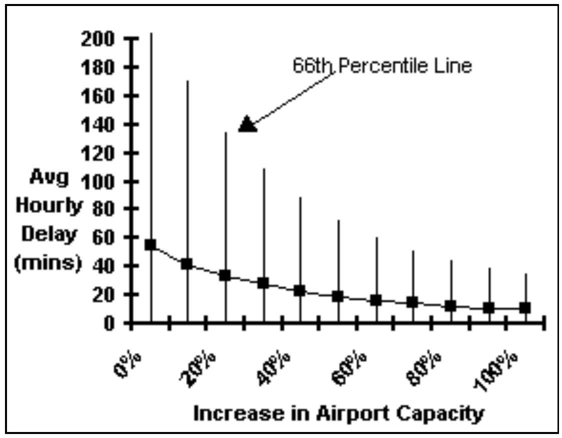

Figure 6. Predictability of Hourly Airport Delays

PREDICTABILITY OF DELAY

Predictability is one metric that has been proposed as a NAS performance measure. Here we shall explore the extent to which delay is predictable. Although delay is, in general, costly and to be avoided, delay which can be predicted with some certainty is preferable to delay which is chaotic and random. Our delay predictability proxy is its standard deviation. If its standard deviation is high, then the delay is difficult to predict. If, on the other hand, it is relatively low, then delay is much easier to predict.

The 66^{th} percentile and the standard deviation of the hourly airport delays were measured as a function of uniform airport capacity increases (similar to the study shown in figure 3). "Hourly airport delay" is defined as the total delay experienced by all flights (arrivals and departures) over a one−hour interval at a particular airport. Hourly airport delays are much higher than the average flight delays measured earlier, because hourly delays are summed over many flights (typically 30−100 flights). There are fifty−eight major airports in this study, and the simulation computes for a twenty−four hour period, so the sample size consists of 58*24=1392 different hourly airport delays.

Hourly airport delays of zero are eliminated from the sample, because they occur during hours of few (or no) flights, and will artificially lower the standard deviation. In effect, this is a conditional probability distribution: the probability distribution of hourly airport delays conditioned on the fact that those delays are nonzero.

Figure 6 shows the mean and 66^{th} percentile of this metric. The 66^{th} percentile is plotted as a proxy for predictability; the higher the line, the less predictable the delay. It is clear that both the average hourly airport delays as well as its 66^{th} percentile are monotonically decreasing as the airport capacities are relaxed. From this chart it would appear that delay becomes more predictable as capacity is increased.

Figure 7 reveals a different story. Here, the coefficient of variation (or CV—the standard deviation divided by the mean) is plotted; a high CV indicates a very scattered (and thus unpredictable) delay distribution, while a low one indicates a narrow, predictable distribution. Figure 7 shows that the CV begins at the very large value of 3, and *increases* slightly as the capacity is increased, finally decreasing to a value of about 2.4. The decrease from 3 to 2.4 is a 20% decrease, which is small compared to the 100% increase in airport capacity. The CV indicates that delay predictability is only marginally improved as airport capacities are increased.

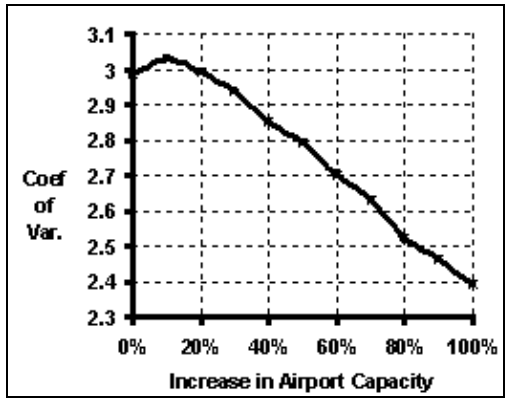

Figure 7. CV as Capacity is increased

Because we have dealt only with aggregate airport delays, it remains to be seen what the predictability results would be for sector or individual flight delays. If the results shown here generalize to other NAS delay measures, then it is quite possible that delay will continue to be a highly unpredictable quantity even as capacity improvements are made.

CONCLUSIONS

It is the consensus of most airline analysts that there exist significant capacity problems in the NAS today, causing measurable, and costly, delays to airlines and general aviation users. Future air traffic growth will exacerbate these problems. We have shown how the DPAT model, seeded with valid data and configured to study specific problems, can be used to estimate congestion−induced delay. The results here would imply that terminal area delay in the CONUS is significant, and will grow in the future. Enroute sector delays, however, are not as problematic. These results are to be interpreted generally across the NAS system−wide; there are, of course, specific, localized areas where the reverse is true. The model also reveals (not surprisingly) that delay is highly unpredictable, but, surprisingly, that delay predictability is only marginally improved as capacity increases are realized.

ACKNOWLEDGEMENTS

We wish to thank Dr. Jonathan Hoffman and Dr. Eric Blair for suggestions for improving this paper. DPAT uses the Georgia Tech Time Warp (GTW) parallel discrete event simulation software. For further information on GTW, contact Dr. Richard Fujimoto, College of Computing, Georgia Institute of Technology, Atlanta, GA 30332−0280, fujimoto@cc.gatech.edu, http : // www .cc .gatech .edu / fac / Richard.Fujimoto

Corporation. For further information, please contact The MITRE Corporation, Contracts Office, 1820 Dolley Madison Blvd., McLean, VA 22102, (703) 883−6000.

REFERENCES

[1] Bruce Nordwell, " ‘Cultural’ Shift Key to New Concept," *Aviation Week and Space Technology*, July 31, 1995, pp. 40−41.

[2] Igor Frolow and Joseph Sinnott, "National Airspace System Demand and Capacity Modeling," *Proceedings of the IEEE*, November 1989, pp. 1618−24.

[3] William Voss and Jonathan Hoffman, "Overview of the ASC System Performance Measurement Project," Proceedings of the First Annual ATM R&D Seminar, Saclay, France, 1997.

[4] R. M. Fujimoto, "Parallel Discrete−Event Simulation," *Communications of the ACM*, 33(10) 31−53, October 1990.

[5] F. Wieland, E. Blair, and T. Zukas, "Parallel Discrete−Event Simulation: A Case Study in Design, Development, and Performance Using SPEEDES," Proceedings of the 9th Workshop on Parallel and Distributed simulation, pp. 139−148, June, 1995.

[6] William J. Swedish, "Upgraded FAA Airfield Capacity Model, Volume 1: Supplemental User’s Guide," MTR−81W16, The MITRE Corporation, McLean, VA; also published as Report No. FAA−EM−81−1 by the U. S. Department of Transportation, FAA, Office of Systems Engineering Management, 1981.

[7] Averill M. Law and W. David Kelton, Simulation Modeling and Analysis, p. 299

[8] Eric L. Blair, et. al., "Comparison of Analytical System−Level Models," MP95W81, The MITRE Corporation, McLean, VA 1995.