Overview of the ASC System Performance Measurement Project

William R. Voss
Federal Aviation Administration, AUA-501
800 Independence Avenue S.W.
Washington, DC 20590 USA

Jonathan Hoffman
Center for Advanced Aviation System Development
The MITRE Corporation, M/S W281
1820 Dolley Madison Blvd, McLean, VA 22102-3481 USA

ABSTRACT

The FAA Office of System Capacity is leading an effort to quantify the performance of the National (USA) Airspace System in terms of its impact on users of the airspace. In coordination with airspace users, these impacts are currently defined in four classes of performance indicators: Flexibility, the ability of the system to respond to changing user needs; Predictability, the variance in the system as experienced by the user; Access, the ability of users to enter the system and obtain services on demand; and Delay, the amount of time over the optimum that it takes to complete an operation. Twenty performance indicators currently exist in these categories, of which many have been quantified into one or more performance metrics. This paper presents results of an initial baseline derived from current operational data and outlines a framework for including these measures of performance in future FAA decisions.

INTRODUCTION

The US Federal Aviation Administration (FAA) program to change its methods of performance measurement is driven by forces from users, from the government, and by advances in technology. Most important of the forces is the growing recognition that users need to control their own operations, that Air Traffic Management (ATM) imposes costs on the users in many ways, and that the FAA should consider these costs as it makes decision regarding the future evolution of the ATM system. To respond to these concerns, there is a need for increased collaboration in tactical decision-making and for the variety of increases in system flexibility that collectively are known as Free Flight. Second, a series of laws and Presidential Orders, most notably the Government Performance and Results Act (GPRA), require increased FAA accountability to users of its services. Finally, a myriad of new technologies offers opportunities for improving the efficiency of ATM specialists. Identifying these opportunities will require improved understanding of the current state of the ATM system.

To be successful, the FAA's new approach to performance measurement had to address several commonly-held beliefs: (1) that the absence of delay represents optimum system performance to the user; (2) other aspects of ATM system performance are not quantifiable, and are best managed through experience and intuition; and (3) the quality and quantity of the data available from the operational system are inadequate to support effective performance measurement. This paper presents some initial results that may eventually cause these beliefs to be abandoned.

BACKGROUND

The traditional metrics for evaluating ATM system performance are throughput and delay, but these measures do not capture the full breadth of impacts that the ATM system has on aviation. Simply counting delay reductions understates the benefits of many critical programs, e.g. conflict probe, and it completely misses the ability of users to select routes and speeds, set arrival priorities, and make other adaptations in response to their own requirements. It is now understood that delay is not always bad: for example, a certain amount of airborne delay at arrival fixes is a sign that no airport arrival capacity is being wasted. Therefore, the scope of performance measurement must be expanded to include, when delay is unavoidable, the benefits that users receive in other areas.

Since late 1995, the Office of System Capacity (ASC) has collaborated with user groups, Air Traffic Services, and future-system developers to create a new, multi-dimensional framework for system performance measurement. This framework has been adopted by the FAA Management Board and industry groups, and work is now in progress to institutionalize key measures by incorporating them into performance plans, investment decisions, and other FAA management efforts.

This framework is "outcome-based" in the language of the GPRA. It does not concern itself with the internal operations of the ATM system, but instead focuses on how the users responded to constraints on their operations. Therefore, there are no performance indicators relating to controller workload, for example, but only how users were affected by the procedures put in place to manage controller workload. Likewise, this effort did not attempt to redefine safety indicators. All of the work that follows assumes all changes in the performance of the system are managed so that the level of safety is continuously maintained or improved.

There are four classes of performance indicators, following the four qualities users desire from the system:

- Flexibility, the ability of the system to adapt to changing user needs;
- Predictability, the extent to which the system permits users to know when they will depart and arrive;
- Access, the ability of the users to enter the system and obtain services on demand;
- Delay, the amount of time over the optimum that it takes to complete an operation.

A "performance indicator", in this context, is a desirable change in the effect that the ATM system has on users. It is expressed qualitatively. The quantification of a performance indicator in a particular domain is a "metric". There can be several metrics for a single indicator. There are, in all, 20 indicators of system performance. Some are quantifiable directly from existing system data, some will become quantifiable as other sources of data are made available, and some are intrinsically qualitative, measurable only by a survey of users of the ATM system. In all cases, the resulting metrics are expected to evolve as users needs or the system architecture change.

A complete list of performance indicators has been published in the *FAA Strategic Plan*[1]. This paper presents selections from the large array of metrics that have been defined. Delay and predictability metrics are frequently two different statistics derived from the same set of data, so they are presented together. Flexibility and access performance measures are presented in their own sections.

PREDICTABILITY AND DELAY METRICS

A recurring problem with delay measurement in the past has been the assignment of causes. When tracking a given flight, it is not always possible to determine that there were, say, eight minutes of taxi time, two minutes of delay due to ATM constraints, and four minutes of delay due to passenger constraints. The delay metrics used here avoid this problem by simply measuring the time it takes to complete an operation, without defining an "unobstructed" movement time as a standard for identifying delayed flights.

Three indicators are addressed here. The FAA will measure success by how much it will

- reduce ground movement times at key airports
- reduce the difference between estimated and actual time en route
- reduce the variation in system performance associated with weather.

"Variation in system performance" comprises several parts; the parts addressed here deal with the uncertainties from day to day in movement times.

Ground Movement

Ground movement information is obtained from the Airline Service Quality Performance (ASQP) database, which is collected by the Department of Transportation for assessing on-time performance of all airlines that carry more than one percent of ticketed passengers in the USA. Times of pushback from gate, wheels-up, wheels-down, and arrival at gate are recorded for all domestic flights for each participating airline, of which there are currently ten.

The delay metric is simply the average time from gate to wheels-up, or from wheels-down to gate, at the 25 busiest airports. The predictability metric is derived by extracting the daily averages for a month, selecting the 75th percentile from among them, and comparing it to the delay metric's value. The result is a time series of monthly values for each airport.

The taxi time data include most delays due to flow management as well as to airport congestion. Many airlines ground crews are evaluated on the basis of how many aircraft push back from the gate at the scheduled time, and air crews are compensated from gate to gate, so there is and incentive to push back as soon as possible, and take ATM delays afterwards. This makes ASQP data particularly useful for this application.

Figures 1 and 2 show the evolution of the metrics for two very different airports since January 1995. One general feature, common to all airports, can be seen in both figures. Average taxi-in times are short compared to taxi-out times, and they are very predictable, generally varying by less than a minute from day to day.

One special feature in both figures can be seen in January, 1996. In this month, exceptionally severe weather struck the eastern US, leading to huge changes from day to day in the availability of entire airports. The effects are seen as far away as Denver; Newark was among the airports that felt the full impact.

Figure 1 shows Denver, Colorado. Denver is a large airport with widely-spaced runways, so between visual and instrument meteorological conditions there is not a great difference in operations. This leads to predictable operations, which is shown in the Figure as a very narrow grey band.

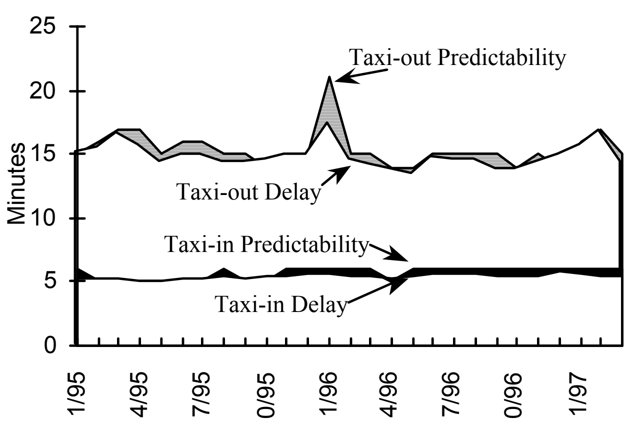


Figure 1. Ground Movement Delay and Predictability for Denver (DEN).

March 1995 is particularly significant for Denver: on the first of the month, Stapleton airport closed, and was replaced by Denver International Airport. Since Stapleton was notoriously susceptible to bad weather, it might be expected that the predictability would improve. However, the first two months of 1995 were unusually mild, so Stapleton operations were very predictable. As the new airport came into operation, average taxi time went up, because the new airport is larger. Predictability was initially harmed, but as airlines and controllers learned the most efficient ways to use the new facility, predictability returned to its old value.

Figure 2 shows a very different story. Newark is a hub for several major air carriers, and its parallel runways are much closer together than those at Denver, which leads to significant differences between visual and instrument operations. Taxi-out times show a cyclic variation with season, longer in summer and winter, shorter in spring and fall.

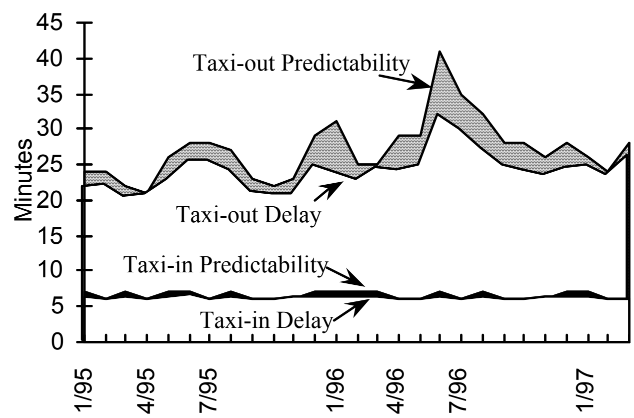


Figure 2. Ground Movement Delay and Predictability for Newark (EWR).

In 1996, thunderstorms were a particular problem at Newark, especially in June and July. The summer unpredictability was even greater than the unpredictability caused by the January snows, mostly because in January, flights were cancelled. A cancelled flight does not enter the delay or predictability metrics.

When the predictability measures of the 25 airports are combined into a single number for each month, a high-level picture of this facet of system performance is the result. Figure 3 shows that both metrics obey the cyclical pattern seen in Figure 2. Also, both delay and predictability grew worse through mid-1996, but have been improving since then.

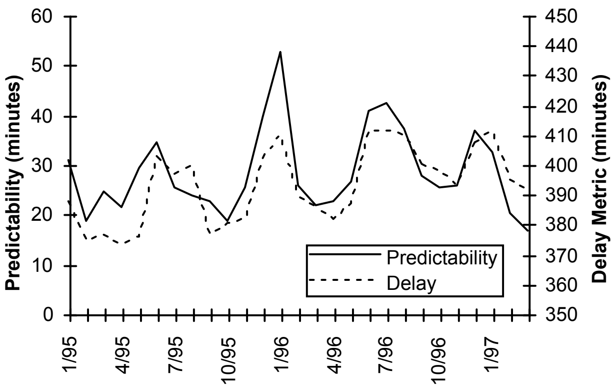


Figure 3. Combined Metrics for the 25 busiest airports: Ground movement delay and predictability

The greatest deficiency in this metric is the incompleteness of the data. Not all airlines are included, and general aviation is totally un-accounted for. It seems reasonable to suppose that ground movement times are similarly distributed among minor carriers as among majors, but we have demonstrated this only at Houston Intercontinental Airport[2].

Airborne Movement

Airborne delay and predictability present particular difficulties. The variation of winds from day to day makes it impractical to simply compare en-route times to each other. To remove the effects of winds, this metrics compares the *expected* time of arrival to the actual time of arrival. In this way, the flight planning systems of the users account for the winds, and only unexpected increases in time are recorded.

The data source for this metric was the Aircraft Situation Display feed to Industry (ASDI). This system provides data from the Enhanced Traffic Management System (ETMS) to users to drive a synoptic view of all IFR flights in the country at a given moment. The estimated time of arrival (ETA) given in this data source is transmitted as the flight enters the en-route system, so the metric will exclude the effect of any departure delay. The actual time of arrival is not available, but a temporary surrogate is the arrival time (AZ time) calculated by the Host computer in the destination Air Route Traffic Control Center (ARTCC). This number is not unbiased; there is a certain offset that has no operational relevance in this context. Changes in the value, however, are significant.

The data are converted to a metric in the same manner as the ground movement. The current archiving of ETMS data at MITRE extends only back to December 1996, so the long time series of the previous subsection are not universally available yet. However, using older archives, an idea of the sensitivity of the metric can be obtained. Figure 4 shows the result of a study of the changes due to the new Denver airport.

Following the bottom edge of the black band, a large variation in delay from month to month can be seen, and predictability from day to day was correspondingly large. When the new airport was opened, the immediate effect was a drop in the mean value of the metric, due primarily to the fact that the new airport is in a different place. Predictability deteriorated for a short time as people familiarized themselves with the new operating conditions, then the new airport showed much improvement over the old. Not only has the day-to-day variation in en-route time diminished, but the seasonal fluctuations have almost disappeared as well. (N.B. the gap at May 1995 is a data outage.)

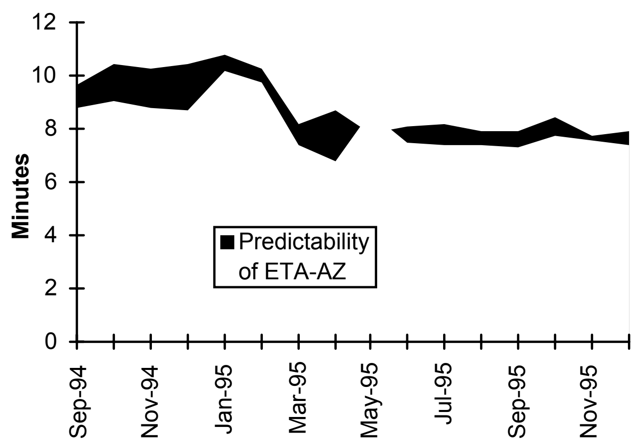


Figure 4. En-route delay and predictability metrics for Denver (DEN). The new airport opened in March 1995.

The en-route delay and predictability metrics show the expected behavior in the presence of large changes in the national airspace system. However, the noise introduced into the measurement by use of the AZ time in place of the actual wheels-down time prevents us from seeing the impact of small changes. Currently, upgrades to the ETMS data feed are being made that will include terminal radar data as well as en-route, which will permit much better estimates of wheels-down time, eliminate the constant bias, and , it is hoped, yield much finer resolution.

FLEXIBILITY METRICS

The *Strategic Plan* states that "FAA will measure its performance by how much it will increase the number of user-preferred routes flown." This indicator has several parts, two of which are analyzed here: preferred ground track and preferred altitude profile.

Ground Track

To identify the user-preferred ground track is not directly possible. Observations of the current system can not permit one to deduce the true user preference, only the user preference in the presence of known constraints. For example, if an IFR Preferred Route exists between two airports, and the controlling Center is known not to grant direct routing, the users will reduce their flight planning workload and simply request the route they know they will be cleared to fly. Many users, in fact, have not even developed the capability to compute a truly unconstrained flight plan. Therefore, this indicator has been quantified indirectly.

Although different classes of users have different preferences, user groups have agreed on this much: that the IFR Preferred Route is not their choice. The metric used, therefore, is the amount of activity off the "pref route", as the IFR preferred route is commonly called. The pref routes affect three phases of flight: scheduling, flight planning, and tactical movement. At each phase, different levels of impact are seen.

Figure 5 shows the difference between great-circle distance and pref route distance for high-altitude IFR preferred routes listed in the National Flight Data Center databases. (Similar metrics for low-altitude and tower-enroute control preferred routes have also been computed[3].) This extra distance has been weighted by the average IFR traffic count between the origin and destination of each route to give final units of aircraft-miles. The figure shows the 20 domestic ARTCCs, colored according to the weighted distance difference. A route that passes through several ARTCCs counts for each of them, since all parties would have to agree to change a route.

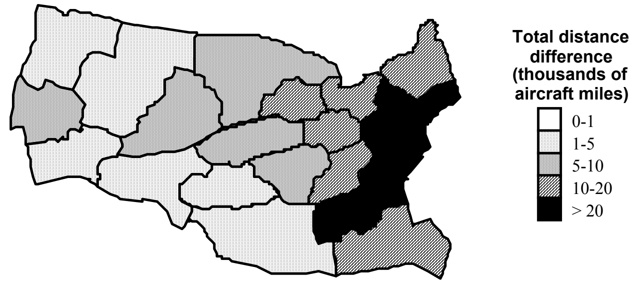
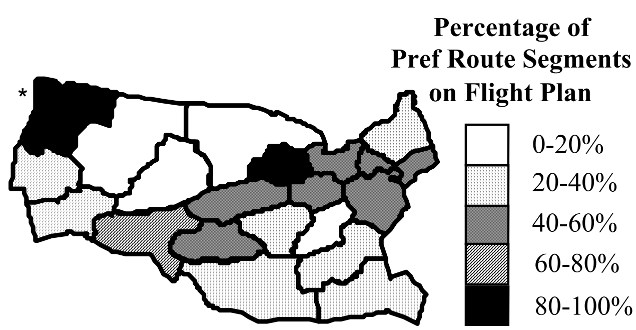


Figure 5. Weighted distance difference between High-Altitude IFR Preferred Routes and great circles.

As might be expected, the heavily-congested (and heavily-structured) eastern corridor shows the most distance difference by far. The western half of the country, where airports are farther apart, shows less impact on users.

To investigate the impact of pref routes (potentially denied user preferences) in the flight planning phase, the flight plans cleared for each flight in the ETMS data were compared to the preferred routes. The first result of this comparison, and the most important number, is that only 28 percent of all traffic flies between cities that are subject to an IFR preferred route. If we restrict our attention to flights that cruise above 18,000 ft., the number increases slightly to 33 percent. The high-altitude figures that follow apply only to that 33 percent of the traffic.



*There are very few flights in ZSE subject to an IFR Pref Route

Figure 6. Difference between cleared flight plan and IFR Preferred Route for high-altitude flights.

Figure 6 shows a comparison of pref routes to flight planned routes for the month of January 1997. To obtain a number that could be geographically resolved, each pref route was broken up into segments between two navigation aids or waypoints. The flight plan is then compared with each segment: if the flight plan passes within 5 nmi of each end of the segment, the flight plan was considered to use that segment. of the pref route.

The ARTCC most restrictive of flight planning was Chicago. Seattle ARTCC (ZSE) restricts 100 percent of the flights in question, but only about 10 flights per day are subject to pref routes in Seattle. Albuquerque ARTCC (between Texas and California) is also very restrictive at this stage of flight operations, due to the presence of a large Special Use Airspace, the White Sands Missile Range.

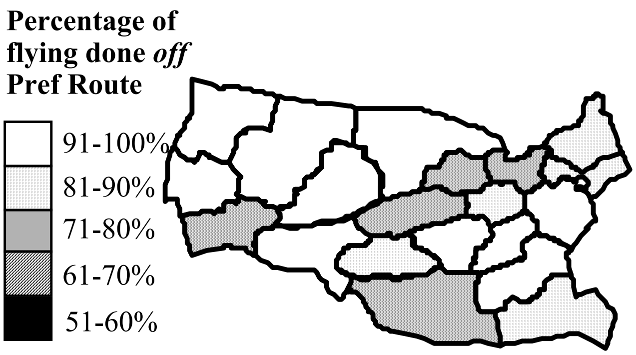


Figure 7. Difference between IFR preferred route and actually-flown trajectory for high-altitude flights.

The last portion of flight operations subject to an IFR preferred route is actual tactical operations. This metric was calculated using the same algorithm as the previous one, with the actual route obtained from ETMS in place of the flight plan. Figure 7 shows this calculation for the same set of flights as Figure 6.

It is immediately clear that the tactical ATM system permits much more activity off the pref route than the flight-planning phase. This metric includes vectoring to maintain separation, re-routing to avoid weather, and direct routes negotiated en route, as well as many other phenomena. Generally across the US, about 40 percent of flights stay within 5 nmi of their flight plan (on average). About 27 percent fly a shorter route than planned, and about 33 percent fly a longer route. These relative values do not seem to change with season, altitude, or length of flight.

Albuquerque ARTCC goes from being one of the more restrictive Centers to one of the most free. This shows the impact of the fact that the White Sands Missile Range is not always in use, and that pilots, FAA controllers, and military controllers can frequently negotiate passage for civilian flights through some portion of the normally-reserved airspace.

These metrics produce values that correspond well with anecdotal evidence about the relative amount of flexibility in the system between ARTCCs, and about the different phases of flight. Their greatest lack is that they are dependent on the NFDC Preferred Route database. There is more structure in the system than that list of formal routings would imply, and the other restrictions are not without costs to users. Agreement on the inclusion of other forms of route restriction has not yet been coordinated with all stakeholders.

Altitude Profile

Another way in which ATM can restrict user flexibility is by assigning inefficient altitudes. It is common in congested areas to segregate traffic to different destinations by altitude. An aircraft that must descend early burns more fuel than one that can cruise longer. To estimate the impact of early descents, we computed the average time from top of descent to touchdown for all flights whose cruise altitude was 29,000 feet or higher.

By restricting the metric to jet aircraft, the user preference for descent time can be modeled very simply. The Jeppesen JetPlan IV flight planning software was used to compute a fuel-optimum descent profile for a large sample of different aircraft. It was observed that all aircraft descending from 29,000 ft. or higher had the same preferred descent time, to within 5 minutes. This five minute interval is the interval between ETMS position reports. Therefore, with the provision that times are not more significant than this, all flights in this altitude range can be grouped together to produce a descent-time metric.

There is currently no single database that can support this metric nationwide, so some modeling approximations were made. For those flights appearing in both the ETMS and the ASQP data, the actual wheels-down time was available. This could be compared to the AZ time, and an average AZ-to-down time computed. This time was added to the time from top of descent to AZ, which was extracted directly from ETMS data.

Results for January 1996 are shown in Figure 8. The airports with mean descent time greater than 41 minutes are shown with black circles; the five airports with mean descent time less than 30 minutes are shown in open circles. (The minimum, fuel-optimum descent time from these altitudes is about 25 minutes.)

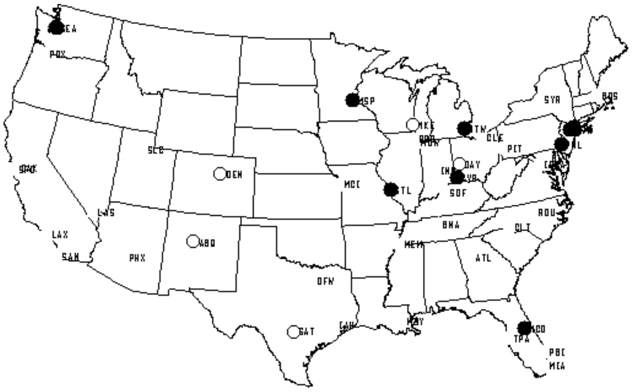


Figure 8. Shortest and longest descent time among major US airports, January 1996

As might be expected, the longest descent times are in New York, where there are many airports close together, and altitude separation may be the only way to segregate traffic. Large descent times are also seen in Seattle (SEA) and Minneapolis/St. Paul (MSP), both of which are far removed from other large airports. In the case of Seattle, the month of January had bad weather that was handled by airborne holding. The holds were below cruise altitude, so the descent time metric includes them. Minneapolis/St. Paul shows no unusual delays, but an examination of the daily totals shows a change in the metric at the beginning of February. This may indicate a change that permits users to stay at altitude longer (Figure 9).

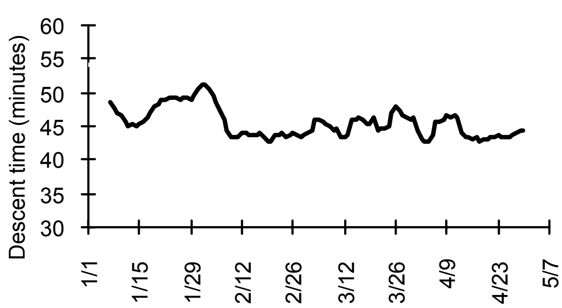


Figure 9. Evolution of descent times at MSP (seven day moving average).

The 5-minute accuracy of the descent time metric in its current, nationwide formulation is not good enough for detailed analyses of local operations. It is also dependent upon ASQP data, which as mentioned above treat only a limited set of flights. This formulation is useful, however, as a guide, identifying areas where local analyses using terminal radar data may be productive. Such analyses are under way.

ACCESS METRICS

As the example of Albuquerque ARTCC above suggests, access to Special Use Airspace (SUA) is of great interest to users. Therefore, the *Strategic Plan* includes the indicator that "FAA will measure success by how much it can increase the civilian utilization of Special Use Airspace."

To quantify this indicator is simple in concept. The number of flights that are seen to enter SUA, expressed as a fraction of those that might wish to. The first number is directly calculable; this effort used the NASPAC Simulation Modeling System Preprocessor to convert ETMS trajectories into SUA entries. To estimate the possible demand for SUA, several days of ETMS data are combined. For each flight that entered SUA, its origin and destination were added to the list of possible city pairs. Then, on a second pass through the data, only those city pairs were kept. In general, the number of flights that might wish to enter SUA is only a small fraction of the total number of flights in the system. SUA tends to lie between direct air routes, not across them.

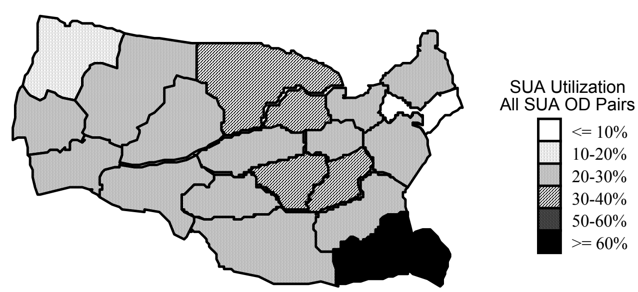


Figure 10. Civilian usage of Special Use Airspace for four sample days in 1995.

Results of the baseline study are shown in Figure 10. Typically, about one flight in four is permitted access to SUA. Exceptions are New York ARTCC, where there is little room to maneuver regardless of SUA, so the metric is low, and Miami ARTCC, in which there is a great deal of SUA off the coasts that lies directly in the path of flights from the Northeast to South Florida. Among the reasons for this SUA is to protect space launches from Kennedy Space Center. This SUA is seldom activated, so civilian traffic frequently gets access, giving Miami ARTCC the highest value of the metric across the USA.

CONCLUSIONS

This work challenges the commonly-held belief that delay is the critical measure of ATM system performance. The metrics presented here show a variety of system changes that would have shown little, if any, impact on existing delay measures. There are opportunities to improve the services the ATM system provides to its users that will not impact overall delays.

This work has shown quantitative assessments of attributes of ATM system performance that in the past could only be described in qualitative terms by participants in the system. ATM service providers have always had a sense of the flexibility they provided to users – these metrics begin to quantify that flexibility. Likewise, users of the system know the level of predictability they can expect – metrics that can quantify that are taking shape as well.

Some aspects of ATM system performance are not quantifiable, and may never be. For these cases, this framework provides for direct input from users of the system regarding their level of satisfaction with the services provided.

The quality and completeness of data used to generate these metrics is always of concern. This work has shown that, even with imperfect data intended for other purposes, it is possible to create measures of performance that indicate the direction of the system as a whole. In some cases the data may represent as little as 1/3 of the traffic in the airspace system, but can provide responsive performance information.

These metrics are part of a common framework that can serve as a basis for comparisons and assessment of impacts of planned actions. These measures are true outcomes in that they are not "owned" by any one individual or organization. Fluctuations in these measures may reflect changes in procedures, equipment, demand, or weather. These measures provide a common vocabulary that allows managers and analysts to make strategic assessments regarding the relative impact of historic events or contemplated actions.

Before these measures can be institutionalized in FAA decision making, three areas of additional work must be addressed. First, the high-level outcome measures described in this paper can not be effectively implemented unless they are connected to a lower-level set of measures that have meaning to the people who operate the system on a daily basis. Executives making choices about the future of the system need outcome measures that provide information about the relative impact of various changes. Managers responsible for the system's operation need output measures that give them information about the effectiveness of their specific actions. Second, it is important to be able to assess a change in a particular measure in economic terms. The dollar value of schedule predictability is known to scheduled carriers, but the impact of changes in flexibility may be more difficult to express. Finally, if it is possible to measure an attribute of the system, it must be possible to model it in hypothetical systems. For many of these metrics, especially delay and predictability, it is simply a matter of re-analyzing output of existing models, but modeling flexibility represents a significant challenge.

Presented at the First USA/EUROCONTROL Conference on Air Traffic Management R&D, June 1997

REFERENCES

[1] Federal Aviation Administration, *1996 FAA Strategic Plan*, Department of Transportation, Washington, DC. http://api.hq.faa.gov/96sp-fin.htm

[2] Hisel, E.P. and J.H. Hoffman, *A Study of Ground Movement Times of Departures from Houston Intercontinental Airport*, MP 96W55, The MITRE Corporation, McLean, VA, February 1997.

[3] Bolczak, C.N., et al., *System Performance Baseline: Initial Indicators, Analyses, and Results*, The MITRE Corporation, WN96W15, McLean, VA, June 1996.