2nd USA/EUROPE AIR TRAFFIC MANAGEMENT R&D SEMINAR Orlando, 1st - 4th December 1998

Use of Performance Metrics in Airspace Systems: US Perspective

Steve Bradford
Federal Aviation Administration, ASD-130
800 Independence Avenue, SW
Washington, DC 20591 USA
202-358-5224, steve.bradford@faa.gov

Lee M. Brown
lbrown@mitre.org

Michelle J. Blucher
mblucher@mitre.org
Center for Advanced Aviation System Development
The MITRE Corporation, M/S W281
1820 Dolley Madison Blvd., McLean, VA 22102-3481 USA

1 Introduction

This paper examines the current use and understanding of metrics in the US Airspace System. It moves us past the definition of metrics covered in the first ATM conference showing the growth in our use of performance measurement and in the increased understanding of the challenges that this places on the Federal Aviation Administration (FAA) and Aviation Community. The challenge, to paraphrase a favorite theme of Carl Schellenberg, the FAA's Assistant Administrator for Financial Services, is to ensure that the measurable does not drive out the immeasurable. To achieve this, the aviation community can not settle for easily measured and casually set objectives. Rather it must establish a context for measurement, use that context to more clearly define system challenges, and set realistic and consistent goals for current and future operations.

We understand that performance measurement has three purposes and stages – the descriptive, the diagnostic, and the prescriptive. The stages of measurement mirror the process by which the FAA and the aviation community are more clearly defining the role of aviation in the larger economy and their individual purpose within that economy. The very act of trying to measure performance has given rise to fundamental shifts in our thinking of the National Airspace System (NAS).

Descriptive measurements and performance indicators give snapshots into the system. The traditional FAA delay metric, daily/hourly operation counts, the number of flights off original flight plans are all descriptive measurements. The great leap forward for the community has been to expand the FAA's vision by understanding that there are other objectives stirring in the minds of the users of the NAS beyond simple delay. The FAA, with the community, has identified these objectives - flexibility, predictability, and access - and has identified performance indicators in each of these areas. Our ability to describe the NAS, and our understanding of the "aviation marketplace" have grown.

The next stage in performance measurement, and the stage the US NAS is now entering, is the diagnostic stage. The diagnostic stage requires that we understand the operational concept, either current or future, and develop our indicators to the point where measurements can be used to identify the "health" of individual NAS components. These requirements have not been addressed. The traditional delay metric is an example of a descriptive measure that is not diagnostic. It describes a performance behavior that reflects a conglomeration of causal factors within which the relationships of the term are not defined, nor the relationship to delay made clear. The delay number is usually insufficient and can not tell whether the NAS has done well given the circumstances, or has under performed badly. There is no relation to "expected" performance.

The final stage is the prescriptive. Once the diagnostic nature of an indicator is shown and the relationship of the indicator to the operational concept and the economic business case of the users identified, it is reasonable to set goals and measure the NAS for significant sustained change in that indicator. Note the caveats – a prescription for an indicator is based on an objective for a change in operational behavior that has a significant impact on the user community. It needs to be able to be clearly defined and unambiguously measured. A goal to improve the peak runway throughput, if clearly articulated, may be a goal for which a throughput indicator can be prescriptive. An n% decrease in the FAA's traditional delay metric is almost surely not.

The other issue that must be addressed in the prescriptive phase is independence. The diagnostic indicators are measurements of aspects of the NAS that often are not independent. In the worst case they are inversely related, as are some aspects of flexibility and predictability, and prescribing large goals to an individual indicator may lead to NAS behaviors which drive other indicators into the "unhealthy" zone.

Finally, the move to increased/expanded measurement requires that the aviation community move quickly from the descriptive to realistic, unambiguous and prescriptive performance indicators and goals. The very act of defining a descriptive measure can lead to goal setting. Whether the goal is unambiguous, under the control of the players, or even set in the right direction need not be clear. In a vacuum, goals will be set, and the FAA will be forced respond. For example, delay reduction, which is a largely descriptive metric, has long been cited as a major objective for the NAS. Using delay as the prescriptive indicator will most likely not support the FAA in determining whether it is meeting its true objectives for modernization. Other new metrics that transcend the descriptive phase must be employed.

2 Approaches to Managing the Performance in an Airspace System

As noted in the introduction, performance measurement is based on a context. Any system can be measured descriptively. The act that moves that indicator into performance is placing it within a context, a concept of operations. We will discuss quickly two approaches for defining current operations. Subsequent sections will discuss briefly how these approaches influence the choice of diagnostic indicators and the establishment of goals. We note that for the purpose of this paper we will over simplify the description of these concepts. There are nuances to both concepts, however, that while these nuances shade the evaluations conducted with the indicators, they do not affect their general selection.

2.2.1 Capacitated Network Management

One approach to an airspace system is to define the system as a capacitated network and manage each resource in the network to ensure, a priori, that the resource does not receive excess demand. In a capacitated network, flight plans are projected onto the resource and the demand for the resources forecast. Any overage in demand is accounted for by either regulating individual paths or scheduled times for resources. The work of the CFMU in Europe, or the Air Traffic Control System Command Center (ATCSCC) during ground delay programs is an example of management of capacitated resources.

2.2.2 Ad Hoc System Management

The approach that the US typically uses is the "ad hoc" approach to resource management. In this environment the traffic builds against the system until a resource is taxed and adjustments are made. Some restrictions are set statically based on worse case – crossing restrictions, preferred routes, etc. Other restrictions are handled more dynamically, set as flow restrictions, for instance, miles-in-trail. Capacities are not established as absolutes, except for serious imbalances due to weather or infrastructure outages. During normal, good weather operations, the NAS is allowed to flow per its own dynamics. The system objective is established to allow for demand to be met efficiently with minimal perturbation to individual flights. All temporary restrictions that effect flows of flights and extend beyond a center boundary are coordinated nationally through the ATCSCC.

3 Performance Measurement

How does one measure performance in the US? How does one tell if the NAS has had a "good day" or a "bad day"? Our traditional method has been to measure delays. For the Air Traffic System this has involved the recording of delays against the flight plan of greater than 15 minutes. This inherently has had the following problem, as noted in the ATS Performance Plan for Fiscal Years 1997-1999 –

> "There are recognized limitations to the methodology currently used by ATS to measure delays in the aviation system. One of the primary problems is that this method does not track any delays less than 15 minutes. ATS records the number of aircraft delayed in excess of 15 minutes during any specific phase of flight. For example, if a flight is delayed during take-off by 16 minutes, that delay is recorded. Similarly, if the same flight is delayed 17 minutes on landing, that delay is also recorded. If the landing delay were only 12 minutes, that delay would not be counted.
>
> ATS is committed to reducing all delays, and is working to define new methods for tracking all delays that are experienced in the system. These methods will incorporate improved information about flight times obtained from aviation users. While such changes may increase the number of delays that are recorded, these improvements to the current delay recording methodology will help ATS in its goal to continue to manage delay and other negative impacts."

The key words are "managing delay and other negative impacts." These words highlight the implicit understanding that there is more to delay than inefficiency and that dynamics exist beyond the FAA's efforts. We describe our efforts to understand, measure, and manage performance delay in both the current and future environments.

3.1 Measuring Performance in the Capacitated Network

Parts of the US national airspace system become capacitated resources only when demand is forecast to significantly exceed the ability of the controllers to handle it. Under normal conditions, this is the case for the airports (DCA, ORD, LGA). Other parts of the system are capacitated for short periods, usually due to inclement weather, equipment outages, or runway closures. These outages and limitations are due most often to inclement weather at or near airports. In such situations the demand and capacity balance is not based on the scheduled demand, rather it reflects an adjusted demand as the carriers cancel flights, delay flights, and in the worst conditions (hurricanes, major east coast blizzards) strategically divert flights to maintain their fleets' operations. In such situations delay is not a good predictor of ATS performance. In fact, delay may be a better judge of how well the community reacts to perceived capacity reductions.

In trying to examine the ability to use delay as a more diagnostic metric, recent analyses have been completed trying to use delay where an individual causal factor has been identified or postulated. In the CAASD analysis of impacts to the aviation users of NAS system outages, methods were used to distinguish delay associated with a service outage from delay associated with weather or other causes. The approach is simple. Days in the month which had no outages are combined to form a baseline. The actual delays on the day of the outage are compared to this baseline to estimate the user impact.[1] If the delays on the outage day are much greater than delays on the baseline day, it is assumed that the excess delay can be attributed to the outage. Figure 1 demonstrates this for an outage on 23 April 1997 at Ronald Reagan Washington National Airport (DCA).

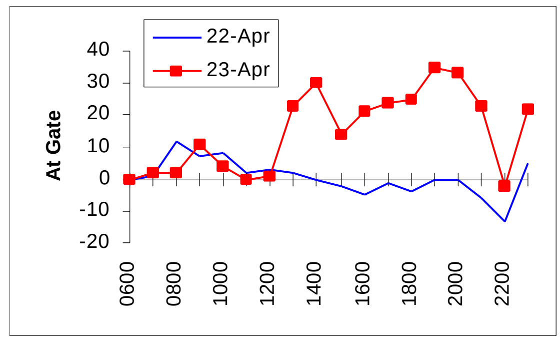

Figure 1. Arrival Delays at DCA, 22 and 23 April 1997

Unexpected events have a clear impact. When warning of a disruption is available, as is usually the case with inclement weather, airlines adapt to the expected conditions.

In another analysis, CAASD looked at the influence of weather on airport disruptions, including delays. The analysis attempted to address the question of "what factor in the weather had the most influence over arrival and taxi delays?" Using the Airline Service Quality Performance (ASQP) data to measure the effect that weather related factors have on
average arrival and taxi delay, the analysis showed that the day-to-day variation that is not weather related dominates the variation that can be directly accounted for by

[1] If weather and outage delays occur at the same time, an attempt is made to find a similar demand day within the month with similar weather conditions, but no system failures.

weather. Figure 2 shows these results with two examples of the data plots used in the analysis.

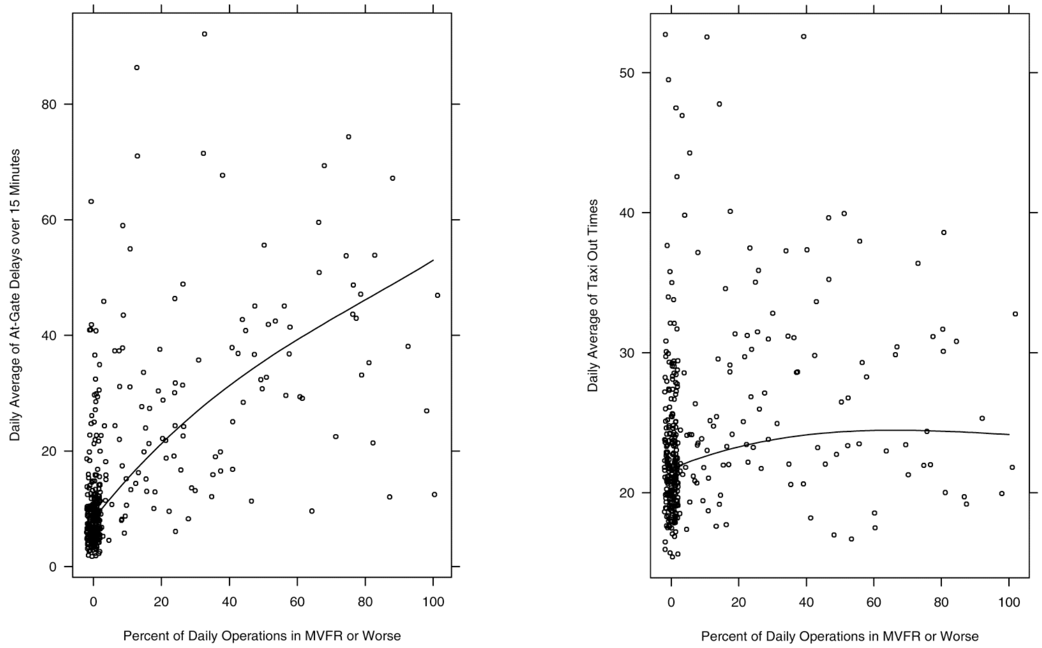

Figure 2. Plots of Daily At-Gate Delay and Taxi-Out Time at Newark in 1996 vs. Worsening Weather

This may seem to be a paradoxical conclusion, given the expectation that weather is the primary cause of delay in the system. There are three causes. First, small-scale thunderstorms far from the airport can cause delays, even when the airport reports visual flight rules. Second, airlines can use their schedule padding to accommodate the weather, which will ameliorate the daily average. Third, and most important, the impact to the major airlines. These users are more likely to decrease demand during weather disruptions through cancellations to maintain their schedules. From the point of view of these major carriers, delays may not necessarily get worse when capacity is reduced (due to weather) because the overall demand at the airport also is reduced.

At a system level, the FAA's ATCSCC is working to develop various utilization metrics that can supplement the standard delay metric to provide better diagnostic capability. These utilization metrics compare actual throughput and demand with throughput targets at key airports. These new measures, being developed with and validated by CAASD, are expected to be useful in identifying situations when the system is working at its capacity, but delays are present due to excess demand. They can also aid in determining when an airport is not fully utilizing its available capacity. In Figures 3 and 4, utilization curves are shown for two like demand days at a busy airport in the US Northeast. In Figure 3, it can clearly be seen that the airport's departure performance degrades rapidly towards the end of the day, showing that the Figure 4 represents a day when the overall utilization and performance for the airport was more acceptable.

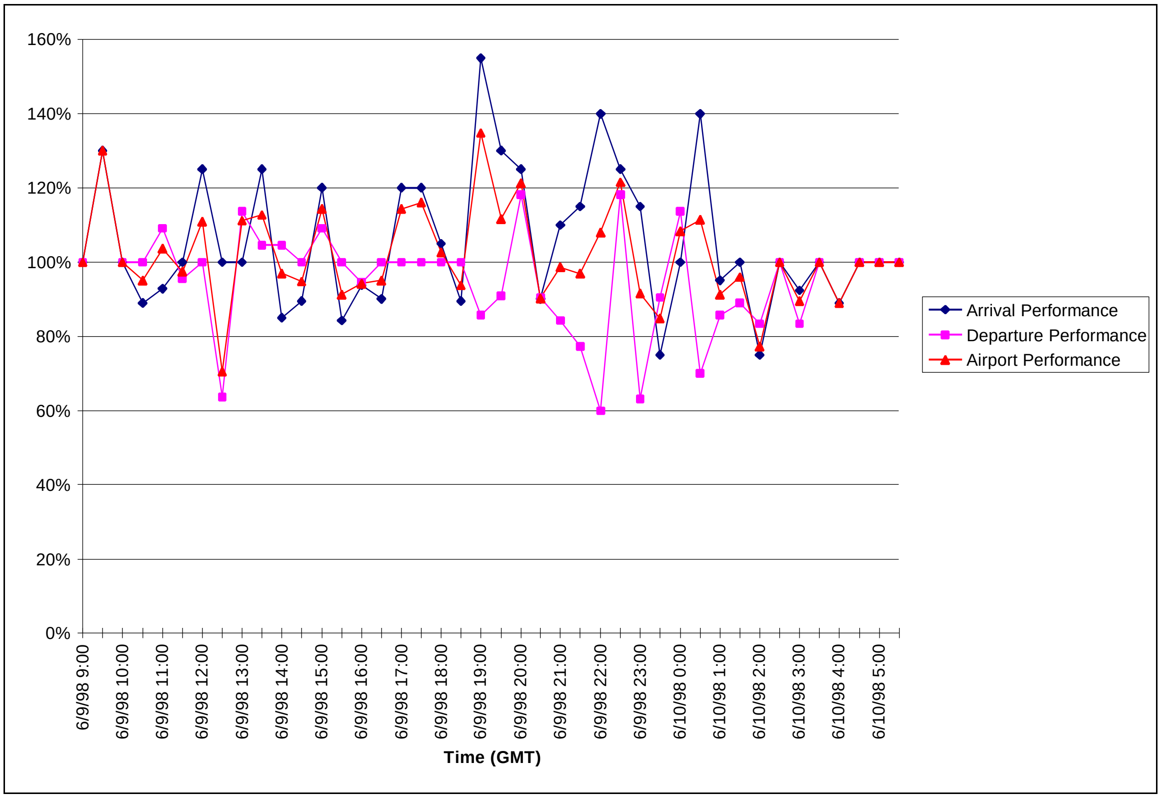

Figure 3. Airport Utilization Curves for 9-10 June 1998

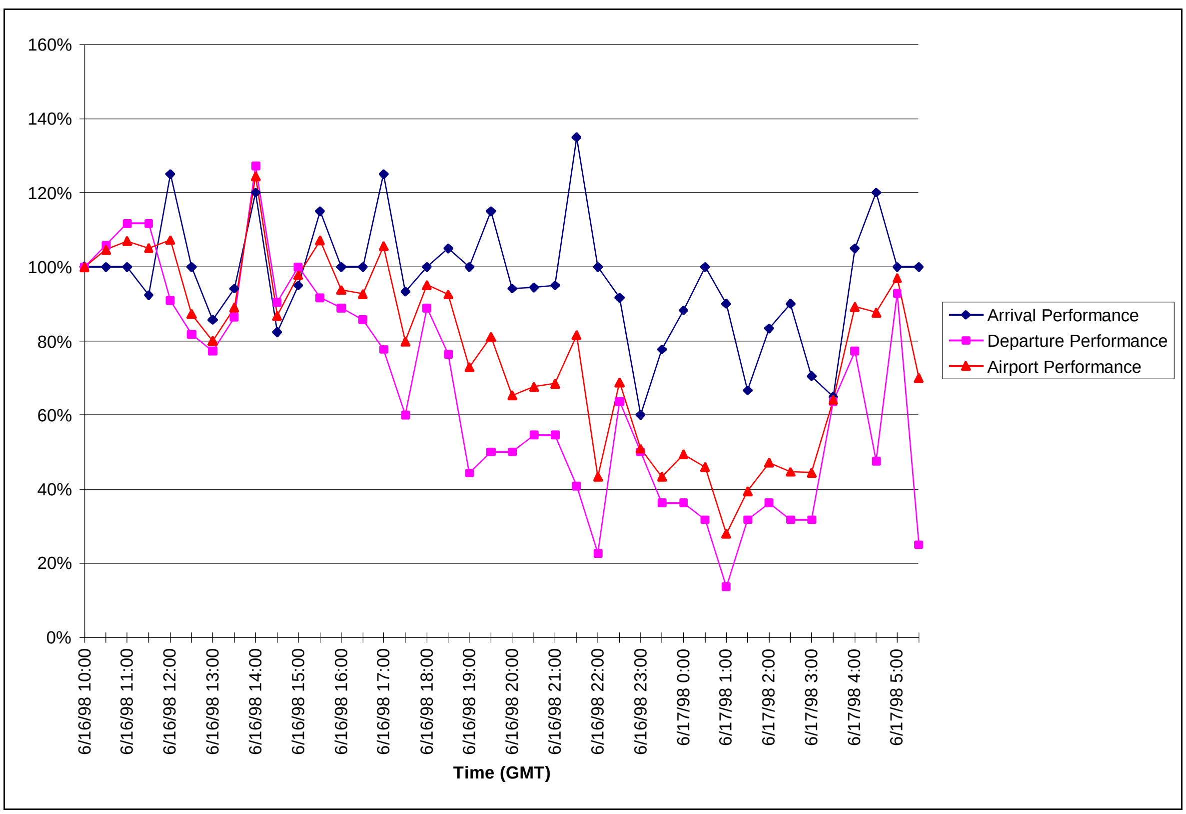

Figure 4. Airport Utilization Curves for 16-17 June 1998

These metrics still do not define the explicit cause for the degradation in performance. But, in the hands of a traffic management specialist with the knowledge of local conditions, they offer a level of diagnostic information that goes beyond delay.

3.2 Measuring Performance in the "Ad Hoc" System

In normal operations, i.e., in the Ad Hoc system, there is no a priori assignment of slots to traffic in the enroute, terminal or airports. Demand/capacity balancing is done in a dynamic fashion. Even in the days when the same dynamic restriction was in place over multiple days, there is no clear trail from identified or potential problem to resolution. In this environment, the approach has been to look at alternatives to examining performance.

One such method has been to measure performance against schedule. A popular way to do this has been to use a NAS-wide queuing tool (NASPAC, DPAT, AND, LMINET) to examine the resource loading and determine both operational delays and schedule delays. While these have been useful in providing additional descriptive indicators of the NAS and can provide insight into its behavior they do not provide a wealth of diagnostic measurement. They are high level representations of NAS processes, not detailed models of the individual factors. When used to replay an existing day, they are often used to provide surrogate data for information not captured or reducible from existing data sources.

The other method is to examine actual behavior against optimized behavior and identify the difference as a delay or inefficiency. This approach is similar to the method used for examining bad weather days and uses the ASQP and Enhanced Traffic Management System (ETMS) data to examine on a flight by flight basis the flight execution against the planned. With this approach a series of descriptive papers have been produced on the performance of the NAS. Most remain descriptive papers, and do not offer diagnostic information because the measurements are at a multi-factor level and do not clearly differentiate between the underlying NAS mechanisms as to the cause of the variation in performance.

Our goal in doing the measurement, analysis and modeling is to not only improve performance within the current system and traffic level, but to also meet the challenges of modernization and the traffic growth. To do this, not only do the measurements need to take place at the multi-factor level but at the individual factor level. Consider the movement of an aircraft through the NAS.

We have methods for capturing samples of time aircraft spend in each of these three phases of flight. This is often the best level of information that we have. We know that on average NAS taxi-out is 3.9 minutes and taxi-in is 1.3. We also know from the work in the ATS Performance Plan that there are wide variations in the taxi-out times and their predictability at the major hubs. At Newark in the period examined the average taxi-out time was in the 23-minute range.

The magnitude of the difference (on a national basis and at specific sites) and the larger degree of unpredictability may lead us to believe that taxi-out performance is a target for improvement. Is taxi-out a problem? What is the cause? Which mechanism? Is it correctable?

Most observations identify the major time activity as the time is spent in queue awaiting departure. Does this make the queue our focus? Only after we understand the component associated with the queue time and the degree to which these components can be affected. If there is a major time component associated with queue management inefficiencies, then we can focus on queue management and procedures systems to improve the queue.

Measuring flexibility is another area of measurement that is being used for diagnostic and prescriptive qualities. We note that the ATS Performance Plan sets a goal of aircraft flying off ATS preferred routes and the mechanism for doing this is the National Route Program (NRP) and the removal of restrictions and ATC structured preferred routes. The NRP provides an opportunity in the flight planning stage to file a non-preferred route flight plan. Figure 5 shows the current metrics being used in the ATS Performance Plan to track increased flexibility. These metrics are still maturing and are being validated by observing their variation as these mechanisms, like reducing ATC preferred routes, are implemented.

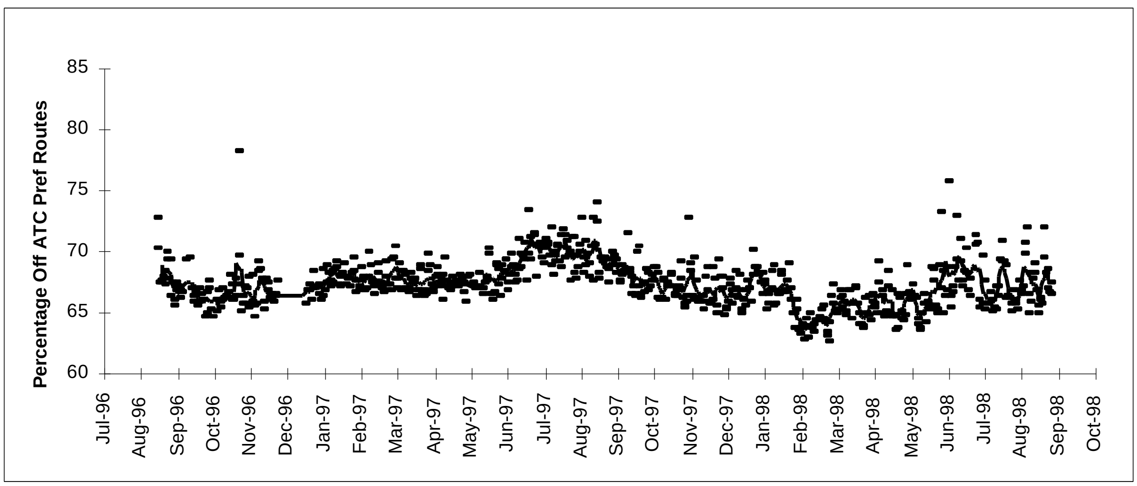

Figure 5. Daily Percentage of Flights Operating Off ATC Preferred Routes

Another objective, for which a goal has not been set, is to allow for more flexible airborne re-routing to adjust for changing conditions and objectives. It is this aspect we wish to look at deeper. One potential indicator is to look at the number of re-routes provided enroute. Of course this is confounded by the fact that the cause of the flight plan change is not captured and could have been initiated by the ATC rather than the controller. Even with this confounding, there is a feeling that more re-routes is an indication of flexibility (perhaps not with enough definition to set a goal). We note that over time there is another factor in the re-reroute total that, if corrected, should drive a portion of the activity and thus the measure in the other direction. We know, anecdotally, that there are flights for which flight plans are filed to get through the automation system and that prior to flight, the first action into the airspace will be to request the intended route. A goal of modernization is to improve the flight planning process both in its ability to support interactive planning, but also to remove the inefficiencies and plan restrictions currently embedded into the logic. If we achieve this objective and understand this mechanism in reroutes, we would expect a drop in that value and should be able to predict that drop.

Flexibility metrics are also being used diagnostically in problem definition activities surrounding airspace design activities. At a growing airport in the US Midwest, flexibility of aviation users to achieve desired altitudes and climb profiles are two metrics used to determine the specific problems with the airspace design around this airport. Figures 6 and 7 show these results.

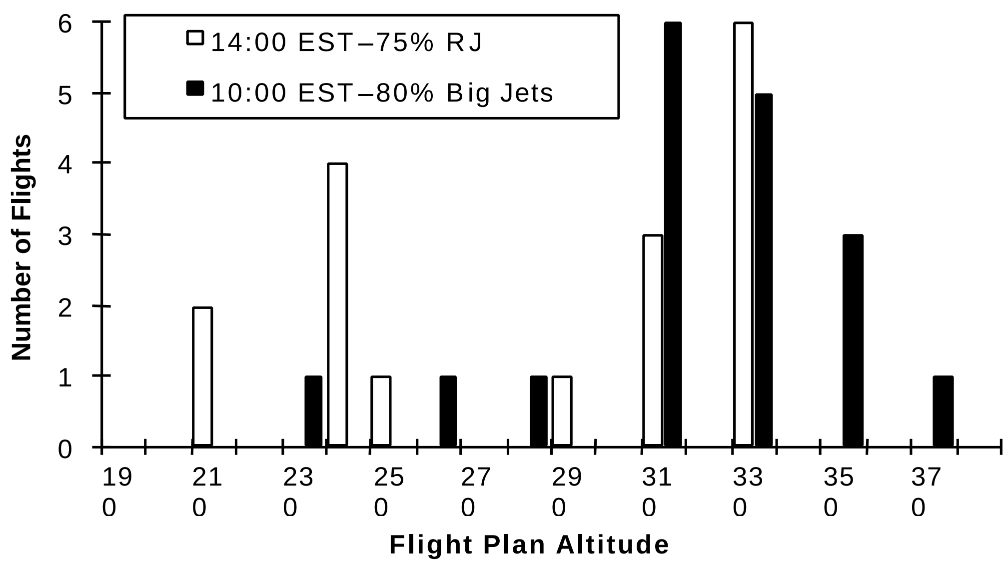

Figure 6. Flight-Plan Altitudes of Jet Departures

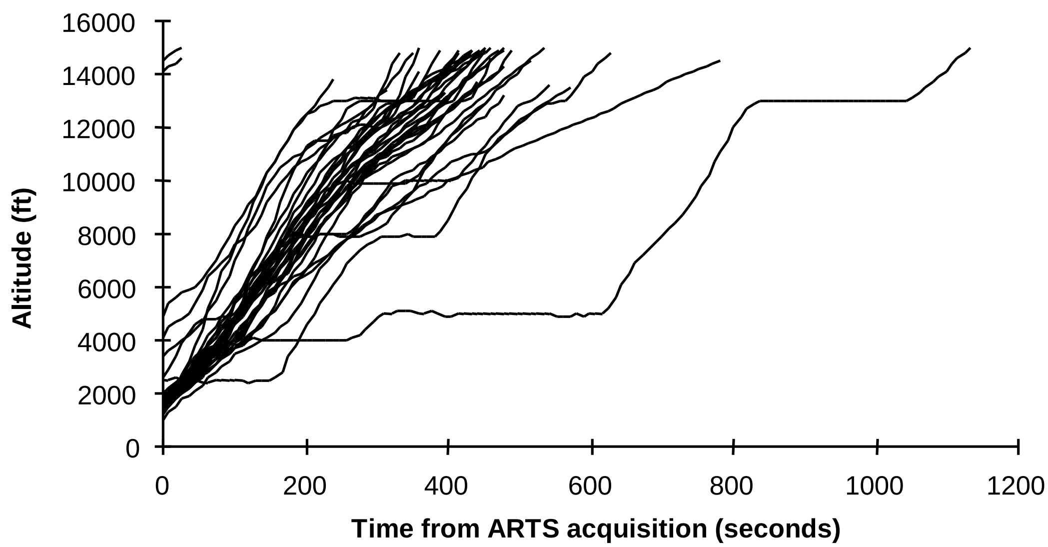

Figure 7. Climb Profiles in TRACON

Finally, it is not the case that all FAA performance measurement is at the same level of maturity. We note that in the area of procedural/design changes the activities clearly indicate all three types of activities. The paper on *Terminal Airspace Performance Metrics – Definition and the Dual CIVET Example* reports on an activity in which indicators track through the descriptive, diagnostic and prescriptive stages. In airspace design, the team gathers data and tries out indicators to build hypothesis of behavior based on the airspace structure and established procedures. In this case, there were inefficiencies in balancing demand to the arrival runways. The descriptive phase continued until the individual factors relating to the problem had been identified and indicators relating directly to these factors chosen. With these indicators and values established, designs were proposed, modeling conducted and the value of the indicators as modeled estimated to choose a design alternative. Estimates of improved behavior were made, and although not clearly established as goals, they do reflect expectations. The CIVET paper highlights the final stages, evaluating the product to gauge the level of improvement achieved.

Conclusion:

The process of performance measurement coupled with the growth of the architecture, the development of the NAS Concept of Operations and the initiation of the National Airspace Review has taken the FAA on a journey from measurement by rote to measurement with thought. As we have increased our quantitative descriptions of the NAS, we have increased our appreciation of the complexity of the NAS. With that appreciation we are moving from a multi-factor appreciation of performance into identifying indicators by which we can better isolate the performance eof individual NAS components and target our improvements. “System engineering” has come to performance measurement.

It is interesting to note that some of the best system engineering practices are exhibited in the development of a non-material solution. As noted in the paper, the examination looked not just at the arrival phase, but examined both the airport and cruise phase to understand the operations and to ensure that the solution was a “system” solution and did not impair any other portion of the NAS.