Performance Measures for Future Architecture

Steve Bradford
Federal Aviation Administration
NAS Concept Development Branch

Dave Knorr
Federal Aviation Administration
Free Flight Phase I Office

Diana Liang
Federal Aviation Administration
NAS Concept Development Branch

Abstract

Aviation growth and the related growing pain are today's headline news. As the world moves toward a more global economy there is an increasing dependency on aviation travel. The Federal Aviation Administration and the aviation community recognize the need to modernize the National Airspace System to support this growth while maintaining a safe and efficient operating environment. To support modernization, new technologies in communication, navigation, surveillance and automation are researched, prototyped and deployed. How effective and efficient are these systems? Are we getting sufficient return on investment? How should these systems be measured? What are the metrics?

Performance measures for the future architecture defines a strategy for breaking up NAS operations into traffic management phases. Each phase has a clear objective need to be achieved and metrics for success. The phases are not fully independent – it has an impact on the adjacent phases. As the FAA fields new systems or implement new procedures to address inefficiencies in the NAS, understanding the targeted phase for improvement will increase the likelihood that the correct metrics has been selected to measure the new procedure or system's effectiveness.

Background

The goal of the Federal Aviation Administration (FAA) is to provide safe, equitable and efficient air traffic services to the users community. To gauge the effectiveness of the delivered services, a set of metrics have been used to measure the performance of the system. Data are collected and results are published periodically.

The historically predominant metrics have been delay and safety. The safety of the National Airspace System (NAS) is based on the number of accidents and incidents in the air and on the airport surface. When considering, system delay increases and decreases are assumed to measure the impact on the users in terms of profit and loss. These are very high level metrics that may indicate trends but may not actually answer any questions regarding the quality of service and the health of the NAS and its users. For instance, a reporting of system delay does not provide insights into where and why the delay occurred. Delays may occur due to weather, airline activities or ATC services, therefore, more detailed information is needed to aid the service provider in identifying and addressing a direct cause and not an easily observed symptom.

As the FAA moves forward with modernizing the NAS, several programs are in place to incrementally implement enhanced capabilities through fielding of new systems. One such program is Free Flight Phase One (FFP1) in which selected sites in the CONUS are scheduled to receive new Air Traffic Management (ATM) systems. How efficient are these systems? Do they deliver the services as forecast? How does it impact the NAS overall? How is it be best measured or measured at all?

The Architecture

To support NAS modernization, the FAA, working with the aviation community has developed a strategic plan (Operational Concepts) for the future of the NAS. The plan outlines the services and capabilities that the agency will provide and the role of the service provider in the future NAS. Along with this strategic plan, a roadmap (Architecture) has also been developed to outline the systems and support activities (people, procedures, training, etc.) that are needed to deliver these capabilities. A schedule and resources are included in the roadmap to explain when the capabilities will be available.

The architecture lays out the capabilities listed in the operational concepts. Each capability is then mapped to a set of systems and the non-systems. Each system and non-system component has a schedule and funding stream associated with it. This traceability allows the decision maker to conduct trades as the FAA is faced with new budget crisis or comprehend the impact to the capability when an individual system slips in schedule. The architecture shows the building blocks needed to provide a capability.

Each capability contributes value to the user. Performance measures are used both to establish the relative value of new capabilities and to measure the actual system impacts. Collecting performance data as new capabilities are implemented is key to future decision making, as implementation priorities in the architecture are refined.

The Metrics

As the FAA proceeds forward with NAS modernization and implementing the future architecture, the agency recognizes the need to modernize and update the metrics that it uses to measure and report its performance. As a result, new metrics such as Predictability, Flexibility, Access, Efficiency, have joined the old stand-bys Delay and Safety. How are these metrics being used? What sort of data needs to be collected? Can we measure real changes in actual performance, and are we doing so? From past experience we know that the NAS is a complex and adaptive system - quickly moving to respond to any changes. With this high degree of adaptation and variability , the agency needs to be vigilant to assure it is measuring itself against the right performance metrics.

For metrics to capture the real impact of NAS Modernization, a more sophisticated approach is required which considers a broader spectrum of measures and a focus on individual phases of flight. The FFP1 program made progress towards an expanded set of metric categories and definitions. These categories map to objectives developed in collaboration with NAS system users and will be used as a reference point for this paper:

Access – ability of users to enter airspace/airports on demand. Runway throughput during peak periods is an example of a specific metric under the access category.

Delay/Efficiency – delay and efficiency are grouped together to capture all changes in mean flight times as well as redistribution of flight times resulting in reduced fuel burn (i.e. spending more time a preferred altitudes). A specific example for Center-TRACON transition is “time from 200 nautical miles to the TRACON meter fix”.

Predictability – focuses on the variation in the ATM system as experienced by the user. Predictability includes both variability in flight times and arrival rates.

Flexibility – focuses on the user’s ability to meet specific flight goals where the above metrics do not capture intent. Included are events where a reduced flying time or direct route is not desired, or when a flight is of such high priority that an air carrier might prefer greater total delays (and increased variability) to allow one critical flight to make connections.

Safety is a difficult value to measure directly since the typical measurements of fatalities or damage are rare events. Safety can be measured by looking at adherence to norms. It can also be measured qualitatively with respect to perception of safety. Since it is both the fact and the perception of safety that we seek to improve in reducing the accident rate, looking for measurement of perception should be an objective of the metrics development work.

Each of these categories contains specific measures for flight segments. Each of these measures are influenced by conditions surrounding the observation. To properly measure or estimate the impact of system enhancements on the NAS we must not only have context information on weather and demand; we need a better understanding of user intent and traffic flow restrictions. Without this information, the measurements discussed above become polluted with data points which sets having completely different objectives and constraints.

Collecting metrics data in concert with weather and demand, user intent, and system restrictions will allow educated trade-offs between individual flight objectives and overall NAS objectives. This approach will also support extrapolation of the benefits of new ATC enhancements to other sites or regions.

How do we establish such a set of data, which addresses user intent, restrictions, and phase of flight?

The Architecture and Performance

Similar to the way in which the architecture is built, performance measures may also be built upon building blocks we will refer to as “traffic management phases”. The traffic management phases are the typical transitional states that individual flights experience in moving through the NAS. It is the management of the individual states objectives that is inherently a part of each controllers job. Controller's do not focus on the movement of each flight and try to optimize each flight, rather the focus is on the process associated with the phase - can they meter aircraft across a fix hitting the spacing required aircraft after aircraft - can they move aircraft out a departure runway with minimal inter-departure time, etc. While the individual flight statistics end-to-end is a measure of the NAS level performance, they currently have not been focused on the individual state objectives and their interactions. A clear understanding of this model of operations and its derivation are essential to attempt measuring the future NAS. Figure 1 depicts a set of phases applicable for gate-to-gate or cruise-to-cruise concept.

What are the parameters, time and space, associated with the traffic management phases? In the US, a typical flight distance is 400 miles and the aircraft crosses two to three centers. The flight runs the gamut from flight planning, ramp, surface, departure, dispersion, cruise, collection, arrival, surface and ramp again. The aircraft departs from the gate, taxis to the runway, takes off, follows a SID, enters the jetway, exits the jetway, queues for the STAR, approaches, land and taxis off the runway to the gate.

There are also flights with distance flown less than 400 miles in the high-density area. For example, flights between Washington, DC (DCA) and Newark, New Jersey (EWR) may never reach cruising altitude. Before the flight can leave DCA for EWR, an arrival slot assignment may be required. The aircraft departs from the gate, taxis to the runway, takes off, follows a SID, queues for a STAR, approaches, lands and taxis off the runway to the gate. For shorter flights, the aircraft will go directly from departure to arrival.

Each flight goes through these phases as it services a certain city pair. Each individual phase impacts its neighbors. An inefficient flow from the collection phase at an airport will impact the arrival phase for that airport. Similarly, it would also impact the previous phases that feed it, such as cruise, dispersion or departure phase. These phases will have to modify their operations because the aircraft they supply can not be accepted by the collection phase. Understanding the relationship of these phases is the first step in measuring the future NAS.

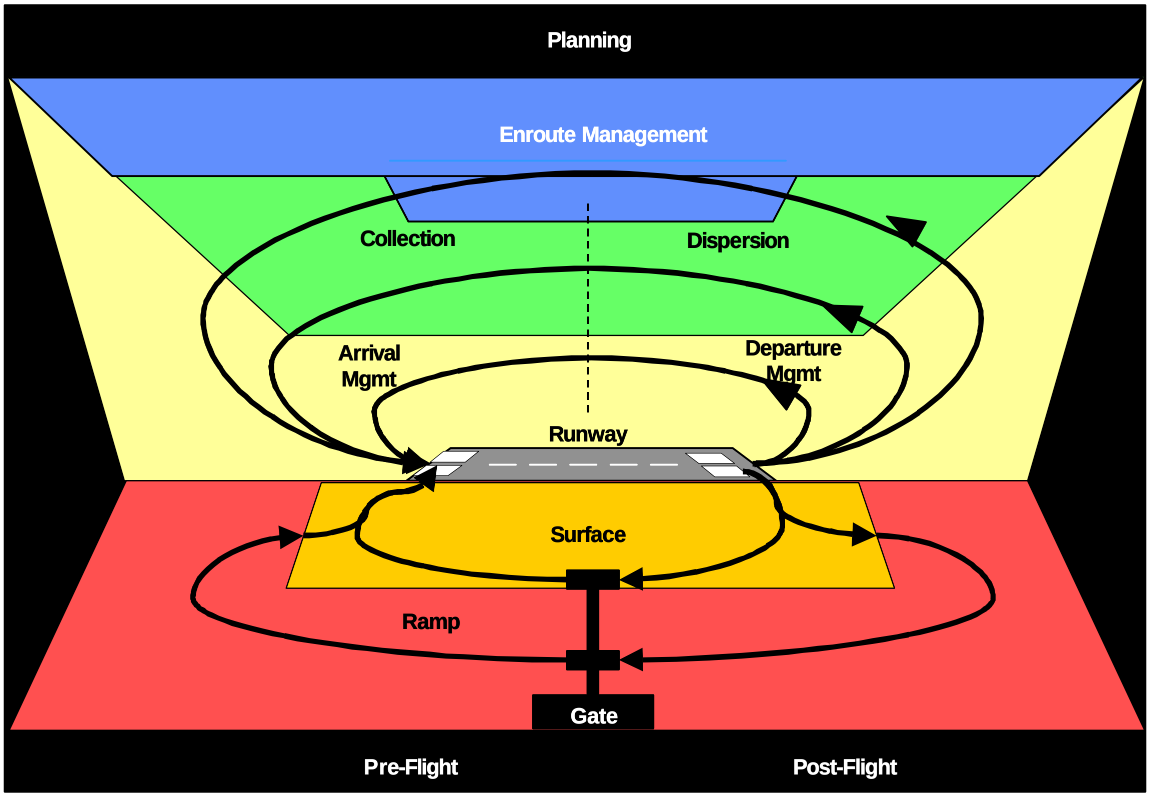

Figure 1: NAS Operation – Traffic Management Phases

The generic NAS operation is now partitioned into smaller phases. The individual phases represented in Figure 1 have a clear objectives and therefore can be measured. For example, what is the right metric for measuring the dispersion phase? Possible metrics are:

- Throughput - does the throughput per time unit of the dispersion phase match the potential throughput of the departure phase?
- Efficiency - can aircraft file their preferred trajectory for this phase
- Efficiency - can aircraft achieve their preferred trajectory in this phase
- Flexibility - is the phase tuned to demand, is the time in phase/extent of the phase tuned to the flow
- Predictability -is the performance across different days and configurations stable.

The comparison of performance also includes a review of services. Is the airspace tuned to needs of this phase, are the Letters of Agreement with respect to this airspace aligned with the flow and operations, are the procedures adequate or is cognitive support required available to dynamically manage the flow most efficiently, do the characteristics of the flow increase complexity for the controller and impact the separation service? Each of these can also be reviewed with respect to the above metrics.

Analysis

Case Study 1 – Free Flight Phase 1 – Measuring Direct Routes in En Route Airspace – Cruise versus Collection Process

Introduction

Free Flight Phase 1 (FFP1) represents a joint FAA/Industry collaboration to accelerate the deployment of new air traffic management capabilities to limited locations in the national airspace system (NAS). The intent of FFP1 is to provide near term benefits to airspace users while collecting data to determine the cost and benefits of a more robust or national deployment of each capability.

One of the key components of the FFP1 initiative is the User Request Evaluation Tool (URET) which supports air traffic controllers in granting user preferred trajectories and/or direct routings. A prototype version of URET is currently operating in the Indianapolis (ZID) and Memphis (ZME) air route traffic control centers (ARTCCs). The FFP1 performance metrics team has been studying the impacts of URET at both of these sites.

URET provides controllers with automatic conflict detection and trial planning capabilities where controllers are able to view alternative aircraft routes and implement those that are most efficient and free of potential conflicts. As of July 1999, the trial planning capability was enhanced to include automated amendments to the Host computer through the URET interface (a “2-way” interface to the Host). The expectation was that automated amendment entries through URET would increase the number of direct routings given by controllers, since URET would support a much less time consuming amendment inputs.

From discussions with controllers at ZID we have learned that all routes through ZID airspace do not equally lend themselves to improvements in direct routings. Some traffic flows are restricted by congestion at destination airports or subsequent en route centers, which may mask overall improvements. To effectively evaluate URET’s operational impact it is imperative to consider the flight “phase”.

Cruise versus Collection State

Measurement of en route performance without consideration of the “traffic management phase” can lead to misinterpretation of new technology impacts and fail to identify priorities in capabilities supporting the future architecture. To properly assess URET’s impact on the air traffic control operation the data set was stratified to identify those aircraft in the “collection phase” from those aircraft in a true “en route phase”. Both states occur with in ZID en route airspace but they are significantly different in their ability to support direct routes.

Direct routing opportunities in the collection phase become much more limited than in the en route phase. By separating aircraft into these two phases and applying metrics appropriate for each we can refine our ability to predict and measure the impact of new capabilities.

Indianapolis Center has some of the most complex airspace in the NAS due to both volume and the amount of crossing traffic. Indianapolis/ZID also has significant constraints on flows heading north and east to busy airports like Chicago, Detroit, Cleveland, Newark, JFK, and Philadelphia. For these airports traffic flows can be set up with miles in trail restrictions going back as far as 400 miles from the final destination airport.

Cleveland (ZOB) and Washington (ZDC) Centers border Indianapolis on the east. They have airspace equally as complex as Indianapolis and have established restrictions on traffic crossing from Indianapolis into their airspace. Centers to the south and west of ZID tend to have less complex airspace with fewer capacity constraints.

Indianapolis controllers have reported an increase in the number of direct routing amendments for aircraft with destinations south and west of Indianapolis Center. We have attempted to validate this anecdotal information by comparing flights within city pairs traveling north from Chicago O’Hare with the same city pairs leaving O’Hare and heading to airports in the south or west (e.g., St. Louis, DFW, or LAX). Sample city pairs chosen for this analysis are presented in Table 1.

Methodology

Several approaches and data sets were considered for measuring URET’s impact on direct routings in ZID. The availability of data associated with the pre- and post-implementation period, however, is a constraining factor. The approach discussed in this paper uses Enhanced Traffic Management System (ETMS) data for two months (October 1998 & 1999) before and after implementation of URET’s “2-way” capability. We have focused this study on Indianapolis Center where controllers have had more exposure to URET and are using the capabilities of the tool more frequently.

South-West	North-EAST
STL	PHL
BNA	CLE
BNA	ORD
BNA	PIT
BNA	BWI
STL	BWI
CLT	ORD
STL	DCA
DFW	PHL
BNA	DTW
PIT	DFW
STL	PIT
IAH	PIT
IAH	DTW
IAH	EWR
STL	EWR
LAX	IAD
SFO	IAD
STL	IAD
LAX	JFK
MCO	ORD
MIA	ORD
RDU	ORD
STL	DTW

Table 1 : Sample City Pairs

To calculate the delta (excess) distance metric we are using x,y coordinates from ETMS to calculate the actual distance flown through ZID (in nautical miles). We then use the entry and exit points of each flight going through the center to calculate the great circle route (assuming it is the optimum distance). We use the difference between the actual distance and the great circle distance to calculate "excess" distance. We are interested in measuring changes in this excess distance associated with the implementation of the two-way Host in URET.

We recognize using distance as a measure of efficiency fails to capture impacts of wind or speed restrictions put on the aircraft. We feel, however, that the impact of wind over the distances flown in ZID airspace would be minimal. Speed restrictions placed on aircraft to achieve spacing and meter feeds to adjacent Centers are not captured with the distance metric.

In its favor, wind generates less variability in a distance metric than it would in a metric like flight time. For this reason, we believe distance is a suitable metric for assessing the impact of URET on direct routes in Indianapolis Center. Furthermore, we believe the great circle route through the ZID entry and exit points is a reasonable measure of "optimal" distance.

Our metric – the difference between the actual distance flown through ZID and the great circle distance (referred to as excess distance) - is intended to capture the difference between the optimal route that airlines *prefer* to fly and the actual route they *do* fly. With ***the hypothesis that URET is contributing to more direct routings through Indianapolis airspace***, a favorable change would identify itself as a reduction in excess flight distance through ZID in 1999 from the 1998 data. The analysis of the entire data set, however, indicates no statistically significant change (1999 from 1998 excess distance).

When sample city pairs are segregated into flows moving north and east (NE) of ZID from those moving south and west (SW) it is clear that the *collection* versus *en route* state phenomena exists (as observed by ZID controllers). Table 2 shows the additional distance required to traverse Indianapolis Center when heading to NE destinations versus the same city pairs with flows to SW destinations. The statistics presented in this table identify a excess distance of approximately .917 nautical miles. This difference is shown to be significant at the α= .05 level. The data used for this test includes the data sets for both 1998 and 1999 and are segregated by NE and SW only.

	NE	SW
Mean	3.761365	2.844028
Variance	33.91213	16.14122
Observations	12156	11996
Hypothesized Mean Difference	0	
df	21613	
t Stat	14.26513	
P(T<=t) one-tail	2.92E-46	
t Critical one-tail	1.644923	
P(T<=t) two-tail	5.84E-46	
t Critical two-tail	1.960075	

Table 2: Excess Flight Distance NE versus SW: t-Test Assuming Unequal Variances

This additional distance for flows NE is an indication that flights heading to capacity constrained airports are, in fact, in the 'collection phase'. Flights heading through ZID airspace to destinations SW exhibit less excess flight distance and are more representative of aircraft in a true "en route" environment. Similar trends exist for traffic heading to Atlanta which is heavily congested (but is actual south of ZID).

URET's Impact: Collection vs. En Route Phases

We are testing the hypothesis that controller's ability to use URET's direct routing capability is measurable when flights are in a true en route phase. Conversely, flights already in the collection phase have constraints, which may override URET's direct routing capability. Initially we analyze the data set going through ZID to the NE.

Table 3 presents the various statistics calculated for NE bound flights in October 1998 versus October 1999. As expected, these statistics show that the average excess distance for NE flights has increased. This difference of nearly a half mile is statistically significant at the $\alpha = .05$ level. Additionally, the observed variance has also increased indicating less predictability in flight distance and time.

	1998	1999
Mean	3.533928	3.97389
Variance	29.67187	37.78622
Observations	5872	6284
Hypothesized Mean Difference	0	
df	12120	
t Stat	-4.182301	
P(T<=t) one-tail	1.45E-05	
t Critical one-tail	1.64498	
P(T<=t) two-tail	2.91E-05	
t Critical two-tail	1.960161	

Table 3: Excess Flight Distance North or East (NE) Bound: t-Test Assuming Unequal Variances

Notably the number of operations in ZID increased by nearly 8% in October of 1999 from the previous year. Additionally, the number of operations at facilities NE of ZID also increased. It is reasonable that these increases in demand where capacity is already limited would override any improvements from URET.

In contrast, Table 4 presents a similar results for SW bound flights for the same two periods. The statistics show that the average delta distance for SW flights has fallen by approximately .316 nautical miles in 1999 versus 1998. Again, this difference is statistically significant at the $\alpha = .05$ level. Variance has also fallen slightly between the two periods.

	1998	1999
Mean	3.00891	2.692897
Variance	18.75015	13.70472
Observations	5737	6259
Hypothesized Mean Difference	0	
df	11334	
t Stat	4.277526	
P(T<=t) one-tail	9.53E-06	
t Critical one-tail	1.644989	
P(T<=t) two-tail	1.91E-05	
t Critical two-tail	1.960175	

Table 4: Excess Flight Distance South or West (SW) Bound: t-Test Assuming Unequal Variances

With the data segregated to focus only on those flights in an en route state we can see the expected impacts of the URET capability (measured in reduced excess distance). Our statistical test alone is not conclusive that URET is responsible for this improvement. Discussions with ZID controllers and data collected on use of the URET tool indicate URET is contributing to an increase in direct routes. As with any statistical analysis expert observation should support the data analysis. Furthermore, we believe the amount of improvement measured may be understated given the increase in ZID traffic from 1998 to 1999.

Analysis of additional ZID data will continue recognizing improvements will likely manifest themselves in flows to the SW where a purer en route environment exists. In the collection phase URET benefits may be related to removal of altitude restrictions as opposed to direct routes. Similar approaches that consider traffic management phases will be used as URET is deployed to subsequent en route centers. As important, traffic management phases must be considered when analyzing potential candidates for future URET deployments and setting expectations for results.

Case Study 2 – Dry Heat Departure – Dispersion Phase

Case Study 2 looks at the performance of the dispersion phase. This example is instructive since the "bottleneck" is not in the much maligned runway or departure phase but rather the ability to get aircraft out of the terminal airspace in a manner, which meets the inherent capacity of the airports. This examples is taken from Airspace Planning activities conducted by the FAA's System Capacity (Office ASC) with the Western Region (AWP).

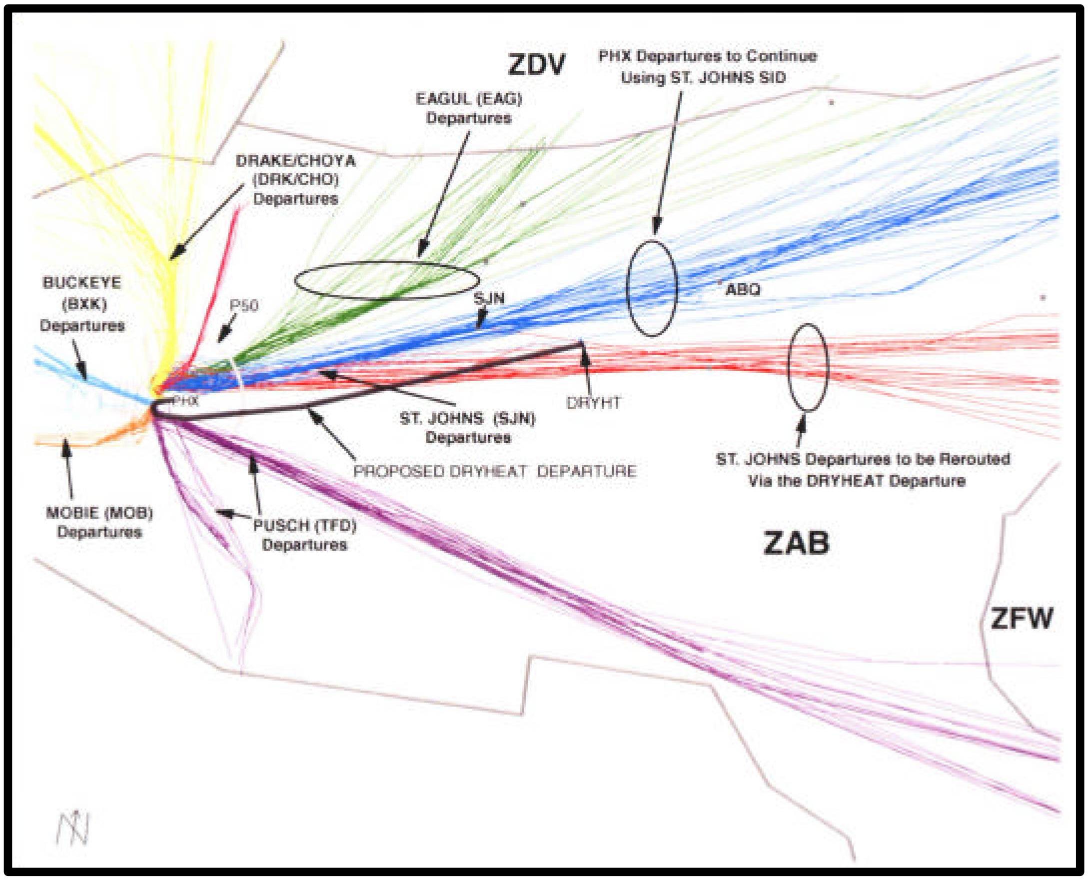

Figure 2 : Propose Dry Heat Departure

Dry Heat is an example where the Dispersion Phase is a hindrance to overall performance. The different management phases had not been balanced to utilize the maximum peak throughput of the airport runways.

The Phoenix International Airport (PHX) is experiencing growth in airport traffic. With the growth, congestion and inefficient operations were beginning to have an impact on the airlines daily operations. As the radar data shows there is significant flow of traffic to the east and northeast from Phoenix.

In the original airspace configuration for the 26R/L runway configuration, the flow to the east was assigned to 26R and departed over the St. John's fix. In this airspace structure the departure fix count for St. John’s was centered around 160 aircraft per day. For all departures on the north runway, the peak day counts were centered around 375 and with the south departure runway counts around 250. The imbalance is clear, so would an airspace redesign for the flow to the east better match the dispersion phase the potential departure throughput?

As part of the redesign effort, the Dry Heat departure (see Figure 2) was proposed and assessed by the FAA and its support contractor. The results reported here are not based on model projection but on a data collection of the area before and after the implementation of Dry Heat departure fix. With the development of a south departure for the east flow, the distribution of departures across the runways is more balanced with the north side peaks near 340 and the south peaks at 305.

The redesign resulted in improvements in both efficiency and predictability. With the Dry Heat departure added, the complexes with significant east flow pushes had reduced total travel times to facility boundaries of 1 minute (20.9 to 19.9) and a reduction in the standard deviation of 1minute. The improved flow is also evident in the taxi-out time numbers. The taxi-out numbers include surface movement and the queue time for departure runway. The decreased values reflect that the departure phase's inherent performance characteristic i.e. potential airport runway throughput, is now match by the airspace structure of the dispersion phase. Average taxi-out times were reduced by an 1.5 minutes (13.5 to 12.0) and have lower standard deviations of 1.9 minutes (7.7 to 5.8).

Conclusion

The NAS is a large, complex system that will continue to change and adapt to new infrastructure enhancement, technologies or procedures. The FAA is working with the aviation community to develop a set of metrics for reporting the health of the NAS and projecting the benefits of improvements. Of equal importance to the improved measures is the identification of not only the operators objective for that flight but also the system's flow objectives at play at various times during the flight.

By looking at the system as a series of interacting traffic management phases, we can partition the end-to-end performance into smaller and more manageable states to assess the problems and proposed solutions, and measure the success of the new implementation (procedures or technologies). For instance, the Free Flight Phase I program is fielding decision support tools to various centers. In some cases the tools will coexist geographically. However, each of the tool focuses on enhancing the operation or objective of a particular management phase in the NAS. By partitioning the flights into the phases and understanding the objectives of the phases, the tool developers can be more effective in identifying the operational need and evaluating their solutions.

Biography

Steve Bradford is the Manager of the NAS Concept Development Branch in the Office of System Architecture and Investment Analysis (ASD). In this role he has participated in the development of the ATS 2005 Operational Concept and the RTCA Joint Government/Industry Operational Concept. His organization is also responsible for leading the effort to validate the future concepts, develop the FAA's ATC Information Architecture and leads co-operative modeling efforts with the European Community via joint agreements with Eurocontrol. Prior to his current position, Mr. Bradford was a team leader in the Investment Analysis and Operations Research Organization where he lead several simulation and analytic software development efforts, and conducted early analysis of Free Flight Concepts. From 1987 to 1991 he worked for CACI, Inc. where he led the SIMMOD model development and taught simulation language and modeling courses. He has also worked for the US Navy developing logistic planning models.

Dave Knorr is team lead for Performance Metrics in the FAA’s Free Flight Phase 1 (FFP1) Program Office (AOZ-40), where he is responsible for estimating the impacts of FFP1 on the National Airspace System. Mr. Knorr holds a B.S. in Mathematics and an M.S. in Engineering Administration, both from the Virginia Polytechnic Institute.

Diana Liang works for the Office of System Architecture and Investment Analysis in the NAS Concept Development Branch. She is responsible for developing Modeling Tools and Fast-Time Simulations to support NAS Operational Concepts. This work includes several models she is developing jointly with NASA and cooperative efforts with Europe via Eurocontrol. Prior to working for ASD, Ms. Liang worked in the Office of Energy and Environment for two years as the lead for the Emissions and Dispersion Modeling System (EDMS), updated the FAA's Air Quality Handbook and reviewed Environmental Impact Statements related to emissions. Ms. Liang holds a BS in Computer Science from George Washington University.