Predicting Congestion in the Northeast U.S.: A Search for Indicators

Emily Beaton
Principal Software Systems Engineer
ebeaton@mitre.org

John F. Brennan
Senior Domain Operations Analyst
jbrennan@mitre.org

James S. DeArmon
Lead Multi-Disciplinary Systems Development Engineer
jdearmon@mitre.org

J. Jeffrey Formosa
Senior Domain Operations Analyst
jformosa@mitre.org

Kerry M. Levin
Principal Multi-Disciplinary Systems Development Engineer
kmlevin@mitre.org

Shane Miller
Senior Software Applications Development Engineer
smiller@mitre.org

Craig Wanke
Principal Simulation Modeling Engineer
cwanke@mitre.org

The MITRE Corporation
Center for Advanced Aviation System Development (CAASD)

Abstract

The northeast U.S. is arguably the most congested airspace in the world. Four major New York airports have very high total operations counts and are concentrated geographically. Improvements are needed for flow managers' decision support systems, to support proactive intervention leading to smoother arrival flows. A CAASD team addressed this issue by investigating predictive "indicators", i.e., quantifications that foretell a future situation with respect to the balance of air traffic demand and capacity at airspace resources. Most flights in the northeast last less than 70 minutes, so predictions of airspace congestion at least one hour ahead would be most useful, since flow control could therefore extend to pre-departure. Predictions are needed especially during visual meteorological conditions, when congestion is not necessarily an expected outcome. Our approach was to examine historical data, in search of identifiable air traffic management problem situations. These situations were then played-back using an integrated real-time model, combining two previously built CAASD systems (the Self-Managed Arrival Resequencing Tool [SMART] and the Collaborative Routing Coordination Tool [CRCT]. The simulation clock was halted one hour prior to the known situation (congested or not), and predictive indicators were evaluated. This paper documents the successful discovery of a congestion prediction indicator.

Introduction

Air travel has become a key element of American life – commerce and leisure activities rely on it heavily. One of the most active markets in the U.S. is the Northeast, more specifically, the New York area, which has become one of the world's busiest air traffic localities. Air traffic management (ATM) of this region is a continuing challenge for the Federal Aviation Administration (FAA). Given the concentration of busy commercial airports (LaGuardia, Newark, Kennedy, and Philadelphia), the airspace requires fine-tuned sectorization to accommodate the significant volume of arrival, departure, and overflight flows [DeArmon (2000)].

Decision support systems for air traffic managers might be improved through greater sophistication in the technology. For this project, we examined predictive "indicators", i.e., quantifications that accurately foretell a future situation with respect to the balance of air traffic demand and capacity at National Airspace System (NAS) resources.

Operational Concept

The FAA's mission need is for a more efficient air traffic system when dealing with congestion, which increases delay and controller workload. Air traffic managers need advance predictions of traffic volume to plan flow strategies and avoid congestion. The airline users also need a more efficient air traffic system, because reducing delay reduces costs due to that delay (e.g., fuel, schedule disruptions). If traffic managers could know, in advance, when potential congestion was going to be great, causing delay, they could act to reduce that congestion, thereby reducing the delay.

In the northeast U.S., where the average flight lasts about an hour, and the ability to modify flights once they have departed is limited, flight planners need to plan changes prior to departure. Predictive indicators with at least a 60-minute look-ahead would provide this needed knowledge of congestion patterns along the route of flight to traffic managers. To be most effective, the predictors would need to be automatically shared between the FAA and the airlines. Each user organization could have its own applications displaying the information in a different manner, but the underlying data should be the same, so every user is looking at the same predictions.

Also, policies and procedures for dealing with the information should be the same for all users across the country. Currently, different airlines work with each respective hub's associated Traffic Management Unit (TMU)s, and procedures might be different from TMU to TMU. With the increasing role of the Air Traffic Control System Command Center (ATCSCC), and increasing collaboration between the FAA and airlines, a standard method of dealing with these problems will need to be developed, comparable to the current method of using flight schedule monitor (FSM) to help plan ground delay program (GDP)s [Metron (2000)]. (TMUs provide flow management services for one of the twenty US Air Route Traffic Control Centers [ARTCCs], these being the facilities for en route ATC. By contrast, the ATCSCC has a centralized role of providing strategic, "big-picture" flow management for the entire country.) For example, airlines could avoid a predicted 30-minute delay by planning a flight to arrive through a different fix. Although approaching from a different direction might normally add 15 minutes to the base flight time, in this case, it may be worth it to save 30 minutes of holding. In another example, an airline, using the predictive indicator, may observe that delaying a departure by 10 minutes could avoid the congestion delay.

Experimental Approach

The goal of the experimental investigation was to identify one or more indicators that were capable of predicting congestion problems. These indicators would be available (1) in time to take preventive action, and (2) in situations in which congestion is not an obvious, expected outcome, e.g., Visual Meteorological Conditions (VMC) airport conditions, or when large convective weather systems are not present. (Flow managers don't need special indicators to know there will be flow problems on an Instrument Meteorological Conditions day).

Therefore, for the experiment, indicators were evaluated according to the following requirements:

- One hour in advance of a detectable problem (defined as an hour during which many aircraft destined for the same airport were subjected to airborne holding),

- For time periods during which all New York area airports were running at or near visual approach capacities, and

- For time periods during which none of the major approach fixes were blocked off due to convective weather.

As described above, congestion problems were defined by the presence of a significant amount of airborne holding. This metric was chosen because (1) airborne holding is a situation that incurs significant delay and cost to the user, and (2) airborne holding can be identified from historical data. Note that some amount of short-duration airborne holding is normal for a busy destination airport, and is often used as a strategy for keeping all arrival slots filled during heavy arrival periods. To avoid defining this situation as a "problem", only days in which at least five aircraft were being held (all of which had arrival times during a single hour) were chosen for study.

Data and Analysis Tools

A set of tools was developed to determine when and where holding occurred for flights arriving at the New York area airports. Several months of data from the Enhanced Traffic Management System (ETMS) were analyzed [Volpe (2000)]. These data include flight plans, periodic (every 4 minutes) position and speed reports, departure messages, and arrival messages. The Holding Analysis Tool (HAT) was developed to scan through daily ETMS archive files and identify holding flights from the position reports. The HAT also computed actual arrival and departure counts for the airports being studied.

For this study, only flights which (1) were destined for EWR, LGA, JFK, or PHL, and (2) reached a cruise altitude of at least 18,000 feet were analyzed. The position history of these flights was examined, seeking groups of frequent heading changes—these are indicative of holding patterns. (This simple technique was validated using visual assessment of some sample cases.) The location, time, and approximate duration of each holding event were stored, along with basic flight information, for further analysis.

These holding event lists were used in two ways. First, the holding events were summarized into hourly counts by arrival airport, and correlated with the arrival/departure counts for those airports. From this summary list, interesting days for further analysis could be selected. Second, interesting holding periods were plotted by HAT for visual analysis (see Figure 1).

The holding plotter and the summary list were used to select 28 situations for study, using recorded ETMS days in January, April, May and June 1999. These were all cases in which the airport was running at near capacity for arrivals, under visual approach conditions. It was assumed that HAT-derived arrival throughputs of about 44 for EWR, 50 for JFK, 50 for PHL, or 36 for LGA indicated that the airport was probably in a VFR operation. The examination concentrated on those hours, and used those figures as the assumed AARs. The capacity of the arrival fixes was set to 15, for all airports. In some cases, there was significant airborne holding, and in others there was no holding at all. The alternative case, Instrument Flight Rules airport conditions and reduced landing rate, was not meaningful for our study, since some airborne holding would be *expected* during bad weather. We needed to find situations where the landing rate was ample, but unanticipated volume alone was the reason for holding. Our predictive indicators, if successful, would help anticipate problems, and let flow management preempt a problem situation.

Modeling Tools

Two software tools were integrated to form the core experiment platform: the Self-Managed Arrival Resequencing Tool (SMART) [MITRE (1999)], and the Collaborative Routing Coordination Tool (CRCT) [Carlson (1998)]. Both tools were developed by MITRE's Center for Advanced Aviation System Development (CAASD) for research sponsored by the FAA. Each tool allows users to view different aspects of the NAS, using either recorded or live data.

SMART was field-tested by cargo carriers United Parcel Service and Federal Express at their operations control centers in Louisville, KY and Memphis, TN. Its success in enabling airlines to prioritize and expedite flights, and in reducing enroute holding, prompted the FAA to request that MITRE transfer the technology to industry. It is now available to be licensed by air carriers, software vendors, and/or service providers.

CRCT has been evaluated locally at the Kansas City Air Route Traffic Control Center (ZKC ARTCC) and is being evaluated nationally at the ATCSCC. In fiscal year 2000 it is planned to be installed at Indianapolis Air Route Traffic Control Center (ZID) as well as ZKC and the ATCSCC, for operational evaluations. It will be available for operational support as needed, but its main purpose at the facilities will be to evaluate coordination and collaboration issues.

<u>SMART</u>

The SMART was originally created as a demonstration project to test the feasibility of airline management of their own arrival flows. The tool allows the airspace user (originally intended to be a flight dispatcher or an Airline Operation Center coordinator) to view the sequencing of aircraft at a resource (the two resources for which SMART has been adapted are the airport and the airport's arrival fixes). The display enables a user to easily see future time periods during which demand at the resource might exceed capacity (see Figures 2 and 3). With this visualization of the relative time ordering of arrival aircraft, SMART allows the user to intervene on a flight-specific basis. The original SMART enabled user intervention by modifying arrival times for both airborne and pre-departure flights. For this experiment, intervention was accomplished using one of two methods: ground delay (or early release) of pre-departure flights at origin airports, or by rerouting flights using CRCT's Trial Planning capability (see below).

The display is a unique (patented) visual representation of a diagonal time line, with deviations from the line indicating expected arrival "bunching" and hence delay. Significant effort went into understanding and refining the data requirements to support the display. For example, SMART was originally designed to display arrivals only at the airports (Figure 2). SMART was modified for this experiment to include the arrival fix displays to visually determine if demand over arrival fixes was going to be a problem even if the airport wasn't saturated. The arrival fix display (Figure 3) is designed to identify flow saturation in the event that fix load balancing could be used to maximize airport throughput.

SMART receives from CRCT the estimated time of arrival (ETA) for all flights bound for the resource. This information enables SMART to create a view of the expected arrival demand.

The x and y axes of the SMART display both measure time, from current time ("now") into the future; "now" is located in the lower right-hand corner of the display. Each open circle making up the diagonal line in the display represents one time slot at the resource. The duration of a time slot is a function of the capacity the user assigns to the resource. Each time slot can be occupied by one flight. Flights are plotted on the graph such that x equals the flight's estimated time at the resource if no other flights were expected, and y equals the flight's assigned time slot at the resource, taking into account other demand. Since a flight's assigned time slot at the resource takes into account all projected demand, when more flights are projected at a resource than there is projected capacity, the time slots assigned to flights become later and later, i.e., the y value of those flights increases. Thus, the flights begin to deviate to the right of the diagonal line. If demand did not exceed capacity, each flight's estimated time at the resource would equal its assigned time slot: x would equal y, and all flights would lie on the diagonal line.

The distance between the location of a flight symbol on the display and its time slot measures the anticipated delay for the flight. For purposes of this experiment, a congestion "problem" is declared when flight delays exceed 10 minutes, and the slots directly afterward are filled.

CRCT

CRCT is envisioned as one component of the traffic management decision support tool set for the 2003-2005 time frame. CRCT may be used to assist users with the following tasks:

- Recognize and analyze traffic flow problem situations
- Evaluate sector loading throughout the NAS to identify possible en route workload problems
- Identify areas where traffic flow needs to be managed
- Identify the aircraft that are planned to operate through the defined areas
- Identify the specific aircraft, that if rerouted, would have the desired effect on the traffic flow management problem
- Define reroutes for a group of aircraft or individual aircraft
- Evaluate the effects of the proposed reroutes on sector loading before the reroutes are implemented

CRCT's Trial Planning function allows the traffic manager to enter a proposed reroute into the CRCT automation. In creating a reroute, the automation calculates a "trial" trajectory, called a trial plan. The trial plan is calculated from the aircraft's current position, through a series of fixes and airways or other route descriptions entered by the user (either typed in explicitly or by using CRCT's graphic rerouting tool), and rejoins the current flight plan route at some point downstream. The user can specify the point where the trial plan rejoins the original route, or the system default can be accepted.

CRCT processes external data consisting of:

- ETMS data (both version 6 and ASDI formats)
- Weather radar (precipitation) data
- Winds aloft and other atmospheric data

CRCT models planned and active flight trajectories and sends predicted airport and arrival fix arrival time estimates to SMART.

The set of CRCT decision support capabilities provides many ways of graphically viewing situational data. Two of the most commonly used displays are the Traffic and Future Traffic Displays for displaying actual and projected flight tracks, and the Sector Count Monitor and NAS Monitor matrix displays, for display of predicted sector workloads.

Integrated Experimental Platform

For purposes of the indicator experiment, SMART and CRCT were integrated so that both displays reflect the same situational information at all times. SMART receives the ETMS data and sends it on to CRCT. CRCT models trajectories for each flight, using the aircraft position data and the latest weather information (when available) to project the ETA for all flights bound for the resource display in SMART. If a user modifies a flight's proposed departure time (through the SMART display), or route (through CRCT's Trial Planning function), the flight's new information (i.e., ETA, arrival fix used, proposed departure time) will be sent to the other system so that it will be consistent on both displays.

Indicators

As described above, the experimental platform allowed formulation and study of several candidate indicators. The initial set included:

- Airport arrival rate [using SMART's airport display]
- Arrival fix crossing rates [using SMART's fix display]
- Close-in and further-out (from the Terminal Radar Approach Control (TRACON)) "ring"-crossing aircraft counts for major arrival flows [using CRCT's flow constrained area (FCA) capability]
- Ring-based wedge-shaped arrival flow rates into the TRACON [also using the FCA capability]
- Sector counts, including red and yellow alerts when thresholds are exceeded [using CRCT's sector count monitors]

Several interesting events were chosen and run through the simulation, with hundreds of different configurations for the indicators. This process revealed much about both the usefulness of the various indicators and the general characteristic of the northeast airspace.

Analysis and Results

The exploration began by looking at all the candidate indicators of future airport arrival flow delay problems. Look-ahead times were also varied, from one half-hour ahead, to three hours ahead, in half-hour increments. Using the ETMS Aircraft Situation Display to Industry feed, confidence in the completeness of the data was observed to be trustworthy up to about an hour to an hour and a half ahead (though that is still very good for decision-making). With the ETMS v6 feed, the look-ahead for carrier flights would be about three hours. So CRCT users with ETMS version 6 may find the indicators useful earlier. The one-hour time period was the most useful filter, because of the ease of correlating the counts with the airport arrival rates (AAR). It was found during exploration that the airport arrival rate and arrival fix crossing rate indicators were the most useful in predicting holding. The other candidate indicators were therefore discarded from further consideration.

Once we had narrowed the possible indicators and determined on a look-ahead time period, we started the experiment proper, testing the indicators on all 28 holding and non-holding events chosen for study. A worksheet was developed to formalize this process. The sheet contained entries for all appropriate configuration data (airport arrival rate, fix arrival rates, dates/times, etc.) The experimenter was asked to evaluate two conditions, noting whether

- The indicator was predicting congestion at the airport or at one or more of the arrival fixes, and
- The indicator was in agreement with the HAT identification of the type and location of the congestion problem (if any)

A matrix (see Table 1) was then used to "score" the prediction. For each event observed, defined as a particular problem time for a particular airport, the results were scored and placed into a database for analysis.

A score of "positive" indicates that the SMART/CRCT arrival flow indicator properly predicted the congestion problem 60 minutes in advance. "False positive" and "false negative" scores indicate erroneous predictions; a false positive was scored if the indicator predicted congestion when there was no holding according to HAT, and a false negative was scored if the indicator predicted no congestion when HAT showed there was holding. In some cases, the results can be ambiguous. For example, if the tool indicates a significant overload at the airport (as defined by at least 10 minutes of delay and slots full via SMART), there should be some airborne holding, and predicting the correct fix is not normally possible or useful. Another situation was a result of our using real historical data, which of course included results of flow actions. For example, in one false positive event, the HAT found no significant holding in one hour, but 60 minutes prior to the hour SMART predicted congestion. When the model continued running past the prediction time it could be seen that the traffic congestion had been dealt with by the imposition of a 20-mile-in-trail restriction. The notes in the table describe how these more subtle conditions were scored.

Two statistical tests were applied to assess our confidence in the predictive indicator. First, we applied a Binomial test, then Fisher's test, using a different null hypothesis for each.

The first null hypothesis is that there is no difference between HAT and SMART—that the difference occurred by chance. Because this research is at the initial stages, a significance level of 5% was used, i.e., we were willing to accept a 1 in 20 chance of rejecting the null hypothesis when we shouldn't.

The Binomial Test was applied, using the *n* tables in Langley (1971). This test is appropriate because:

- There are two major classes (congestion, no congestion),
- The average in the smaller class (HAT, no congestion) is between 10% to 49% (it is 46%),
- The number in the smaller class is between 5 to 20 (it is 13), and
- The sample number (SMART = 14) is greater than expected (HAT = 13).

The observed significance level was substantially greater than the *alpha* of 5%, probably closer to 15%. We therefore cannot reject our null hypothesis that there is no difference between "truth" (as represented by the HAT) and the prediction indicator (SMART).

However, the Binomial Test doesn't take into account the information collected regarding correct and incorrect predictions (the false positive and false negative scores). Of the 14 predictions of congestion, the SMART indicator was correct (i.e., agreed with the HAT) 12 times, and was incorrect 2 times. The SMART indicator predicted no congestion 14 times also, and was correct 11 times and incorrect 3 times.

Langley's *d* tables for Fisher's Test were used to examine the probability that these results could have occurred by chance, if there was no difference between the two matched observations. Fisher's test is similar to Yate's χ^2 test, but applies to small numbers of observations, where N is 8 to 50 (in this experiment, N = 28 total event observations). It is used to compare differences in proportions.

For Fisher's Test, the sense of the null hypothesis was switched. It is a stronger conclusion to be able to reject the null hypothesis, rather than conclude, as with the Binomial Test, that we cannot reject it. Therefore, our second null hypothesis is that the SMART predictions reflect pure chance outcomes—which any random predictor (e.g., coin flips) would perform just as well. Once again, a significance level of 5% was used.

The observed significance level in this case was less than 1%. This means that we must reject our null hypothesis—the probability of these proportions of congestion predictions happening merely by chance is less than 1 in 100.

Conclusion

The SMART display, at the airport and the arrival fixes, was found to be a useful indicator of future problems. Therefore, there exists an indicator, which can predict air traffic congestion into the Northeast airports, in situations in which congestion is not an obvious, expected outcome (e.g., VMC conditions at the airport). This indicator provides a one-hour forecast—enough time for traffic flow managers or airline operations personnel to take preventive action. The information provided could help flow managers determine the best management technique for dealing with the problem. Sharing the information with airline operations could enhance collaboration, increasing the options considered for dealing with the problem and improving further the flow decisions made. The SMART display may be useful for any merge points, not just arrival fixes and airports.

Although we believe we have demonstrated this to be a plausible concept, more study is needed before testing in the field. Incorporating additional information into the display would enhance its capabilities (e.g., illustrating by color-coding the actual minutes of delay, and providing actual numeric flight demand per user-specified period); also the indicator needs further validation to better understand under what conditions the indicator does and does not work. A relatively small sample set was used; still needed is to determine what the full sample set should be, and complete the research with that set. Also, a problem was defined as a group of flights culminating with at least a 10-minute delay, and no open slots surrounding the group. More research is needed to refine that delay definition.

Once all these improvements have been tested in the lab, the indicator needs to be tested operationally, in the field. At this time the optimum users could be determined; whether they are the ATCSCC, the ARTCC TMU, the airlines, or all of them.

References

Carlson, Laurel S. et al, 1998, *Operational Concept for Traffic Management Collaborative Routing Coordination Tools*, MP98W0106, The MITRE Corporation, McLean, VA

DeArmon, J. S. et al, 2000, “Complex, Congested Airspace and Impacts on Airline Operations,” *Handbook of Airline Operations, First Edition*, McGraw-Hill, Washington, DC

Langley, R., 1971, *Practical Statistics Simply Explained, Revised Edition*, Dover Publications, New York, NY

Metron, Inc., 2000, www.metsci.com

MITRE, 1999, *MITRE’s MSR/MOIE Spotlight, The Self-Managed Arrival Resequencing Tool MOIE*, www.mitre.org/news/articles_99/msr.shtml

Volpe National Transportation Systems Center, 2000, www.volpe.dot.gov

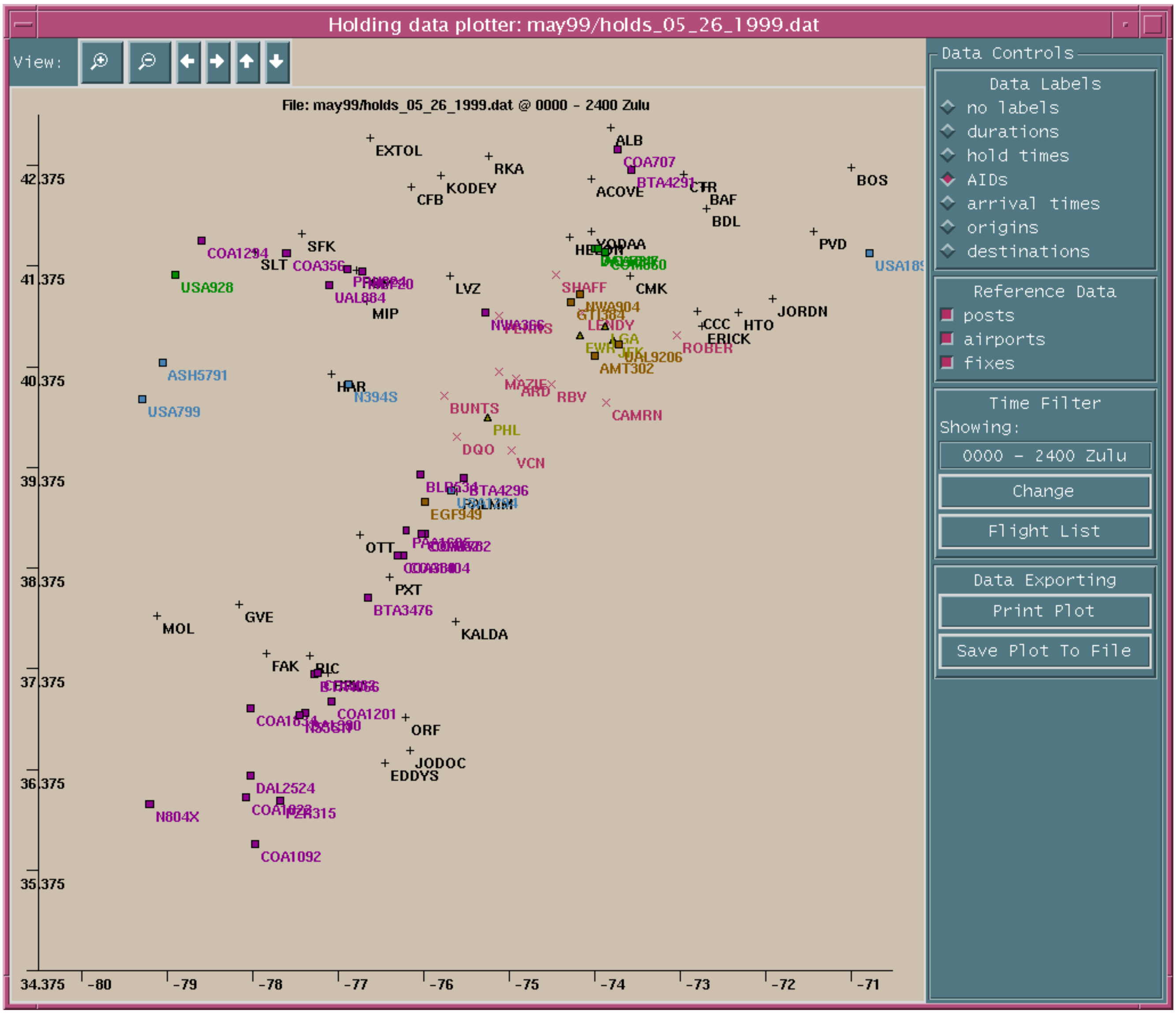

Figure 1. Holding Analysis Tool (HAT) Plot

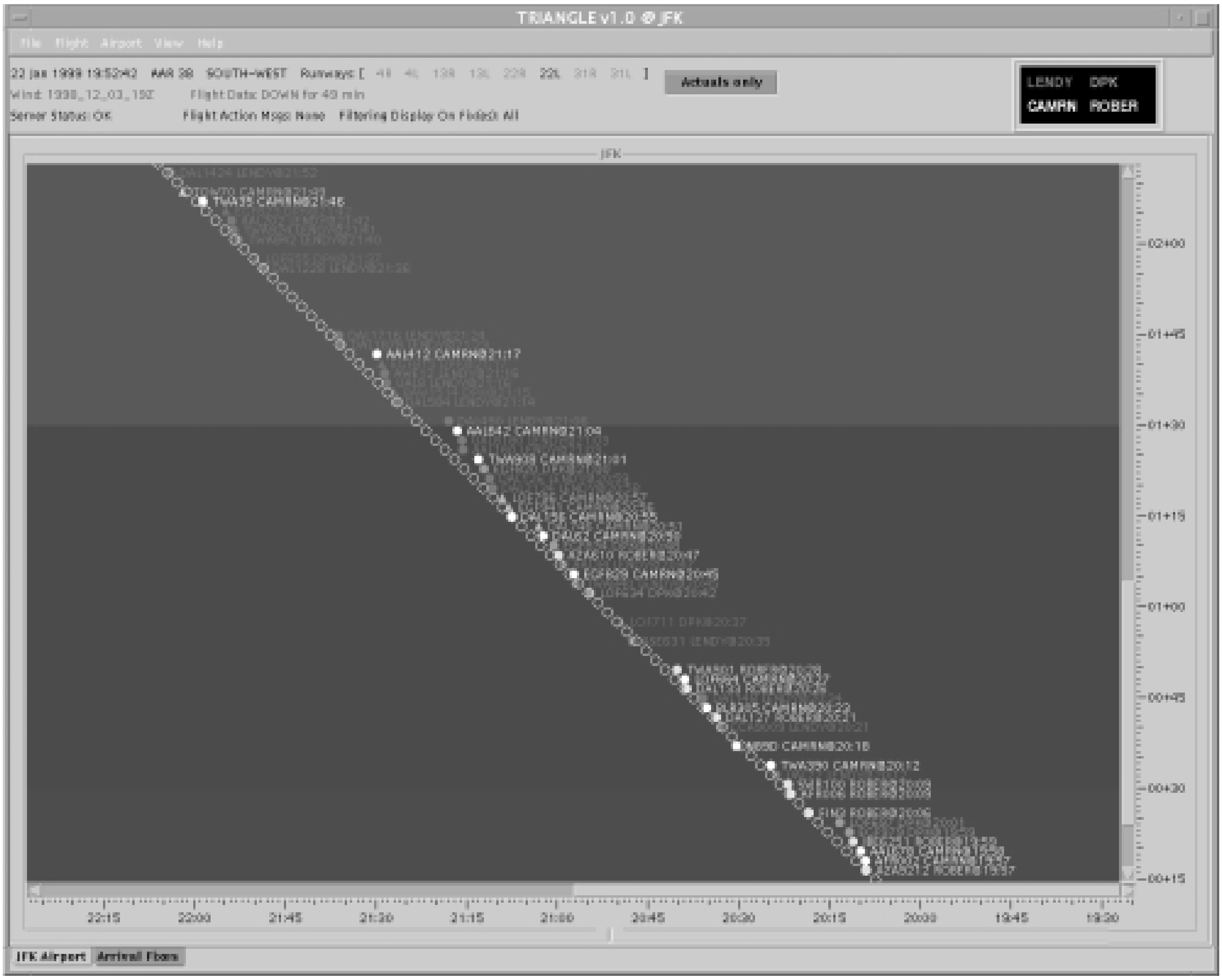

Figure 2. SMART Display at Airport

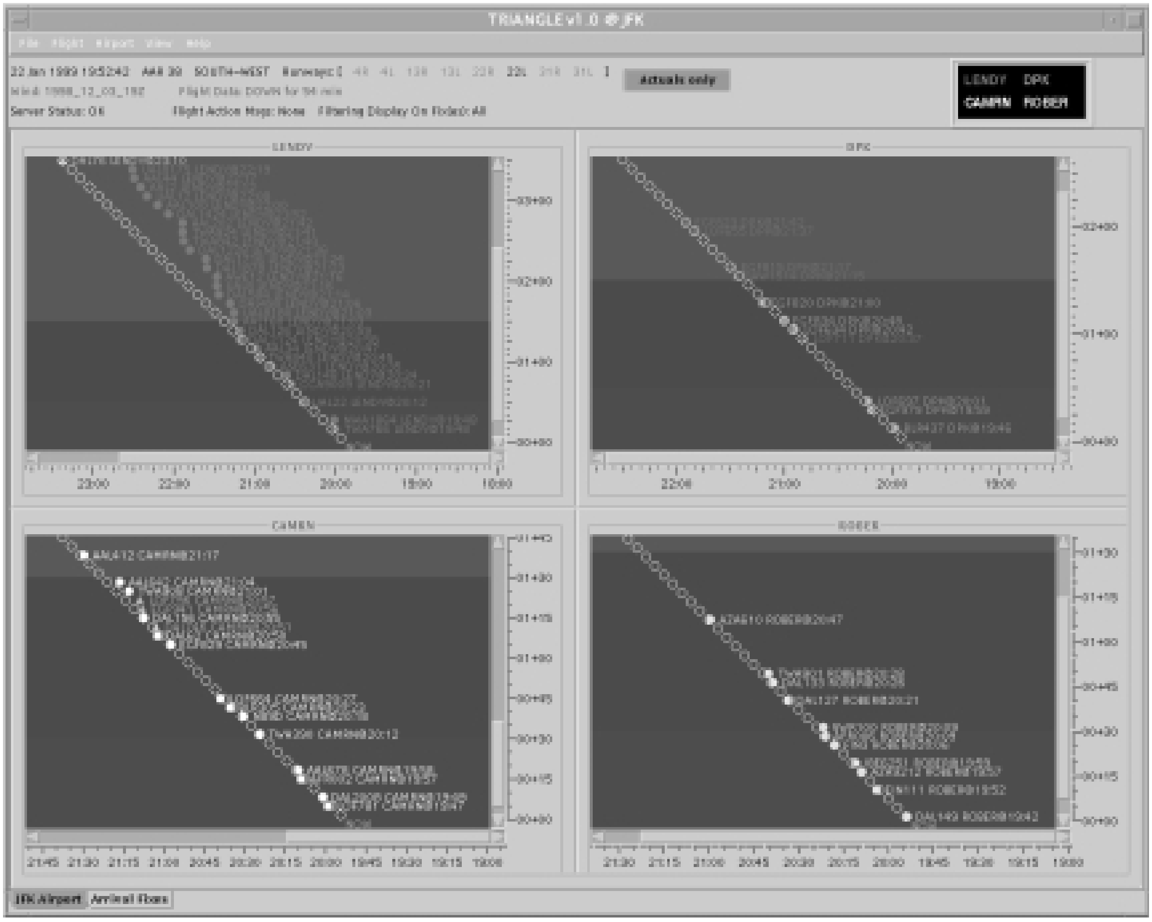

Figure 3. SMART Display at Airport's Four Arrival Fixes

Table 1. Prediction Indicator Experiment Scoring Results Summary

Holding Analysis Tool (HAT)		**Prediction made by SMART arrival flow predictive indicator tool, 60 minutes prior to congestion problem**				
Problem Class	Number of Events	No excessive demand	Excessive Demand on One Arrival Flow		Excessive demand on multiple arrival flows	Excessive demand on the airport (VFR arrival capacity)
			On HAT arrival flow	On different arrival flow		
No significant holding (< 5 ac)	13	Positive 11	Not applicable	False Positive 2	False Positive 0	False Positive 0
Significant holding (>= 5 ac), one arrival flow	13	False Negative 3	Positive 5	Ambiguous See Note 1 1	See Note 2 0	Positive 4
Significant holding (>= 5 ac), multiple arrival flows	1	False Negative 0	Positive 0	See Note 2 0	Positive 0	Positive 1
Significant total holding, but not >5 at any one flow.	1	False Negative 0	Not applicable	See Note 3 0	See Note 3 0	Positive 1

Note 1: Problem is predicted at a different fix, but this may be an appropriate flow solution to the congestion problem. Anecdotal evidence indicates that flow managers occasionally hold a light flow to allow a heavy flow to “clear”.

Note 2: If there is NOT excessive demand on the airport also, then this result is scored as ambiguous. Otherwise it is a positive result.

Note 3: If there is NOT excessive demand on the airport also, then this result is scored as a false positive. Otherwise it is a positive result.