Predictability and Uncertainty in Air Traffic Flow Management

Joshua W. Pepper, Kristine R. Mills, Leonard A. Wojcik
Center for Advanced Aviation System Development (CAASD)
The MITRE Corporation • McLean, Virginia 22102, U.S.A.
Jpepper@mitre.org, kmills@mitre.org, lwojcik@mitre.org

Abstract

This paper presents an analysis of traffic flow management (TFM) events of two types: en route events in the Pennsylvania (PA) region of the U.S. and events affecting the Chicago O'Hare airport (ORD) terminal area. We present a method of accounting for uncertain weather information at the time of TFM decisions, based on Bayesian decision networks. However, we show that data from past TFM events is, by itself, insufficient to distinguish between the efficacy of different strategic TFM decisions, at least for delay, cancellation, diversion, and departure backlog performance metrics. Patterns in TFM performance metrics exist, but there is wide variability across TFM events. Other, less comprehensive metrics that address how well TFM plans execute without undesirable modifications may distinguish among TFM actions better. Modeling as a means to augment data from actual TFM events is discussed. Learning and adaptation implications for the TFM system are presented.

1.0 Introduction

In the U.S., airline schedules are challenging even on good weather days. When bad weather limits the capacities of airports and airspace, U.S. Federal Aviation Administration (FAA) TFM specialists at the Air Traffic Control System Command Center (ATCSCC) may institute various TFM initiatives to manage excess demand. These actions are undertaken as part of a collaborative decision-making (CDM) process involving the FAA and major airlines. FAA TFM actions can be divided into strategic actions, which are typically taken at least 2 hours before weather is expected to affect operations, and tactical actions, which are taken within 2 hours of the weather.

Strategic TFM actions include ground delay programs (GDPs), which reduce the demand to a given airport by spreading out the original schedule over time, and "playbook" actions, which reroute large blocks of traffic around regions of en route airspace according to predefined plans. Tactical TFM actions include ground stops (GSs), which stop flights on the ground that are due to arrive later at a given airport, coded departure reroutes (CDRs), which are reroutes for specific flights from a given airport, and reroutes around the weather. In addition, airlines may respond to weather or forecasted weather with flight cancellations and, if necessary, diversions.

Strategic TFM decision making may take place in the context of significant uncertainty with respect to both demand and weather information. With the advent of CDM, there has been an overall improvement in the extent and quality of information exchanged regarding departure times and cancellations [1]. This paper does not specifically investigate demand uncertainty, but rather focuses on the effects of weather uncertainty. Weather forecasts may convey only a likelihood of weather problems, and may be uncertain in terms of intensity, location, time of onset, and duration. The capacities of National Airspace System (NAS) resources, including airports and airspace, depend critically on the nature and extent of weather problems.

A key to improving TFM is to understand how to account for uncertainty in the demand and capacity of NAS resources. Recent analytic work on the demand side includes a study of GDPs simulated as single-server queuing systems, with demand uncertainty approximated in the form of cancellations, unexpected arrivals, and aircraft arrival time drift [2]. In operational decision making, a feedback process is needed to be able to learn from past experience, and the feedback must account for the fact that decisions are made with uncertain information. At ATM 2000 and 2001, we presented research on the application of agent-based modeling to understand TFM decision making with certain and uncertain weather information [3, 4]. This paper presents an analysis of uncertainty in weather predictions across two types of events in actual system operations during years 2000 to 2002, and relates this operational experience to the decision analysis approach presented at ATM 2001.

2.0 Types of TFM Events Considered in This Analysis

2.1 En Route Weather Events in the PA Region

En route airspace in the PA region is heavily traveled by flights to and from airports in the northeast U.S. Traffic to and from these airports, which include New York LaGuardia and JFK, Newark, and Philadelphia airports, has key economic and operational importance in the NAS. Under reduced en route capacity, airborne eastbound flights to the northeast airports have priority over outgoing westbound flights waiting on the ground. Arrivals may begin to deviate onto the departure routes, which limits the ability of the northeast airports to depart aircraft and leads to departure backlogs. These departure backlogs cause operational problems as large numbers of aircraft impair ground movement.

When convective weather in the PA en route region is forecast, FAA can respond strategically with GDPs into the northeast airports as a means of limiting en route airspace demand. These GDPs are described as being in support of the Severe Weather Avoidance Plan (SWAP). Other strategic options include transcontinental playbook reroutes around the affected en route airspace, or simply waiting until the weather situation is clearer. As the scenario progresses, FAA's tactical options include GSs into the northeast airports, CDRs from the northeast airports, and smaller tactical reroutes.

Weather forecast information is from the Collaborative Convective Forecast Product (CCFP), which produces forecasts at 2-hour intervals. (However, during data collection for this study they were produced at 4-hour intervals.) The CCFP forecasts areas of convective activity. Associated with each area are coverage and probability (Table 1). These forecasts are challenging for TFM operations because their uncertainty is so great.

The analysis presented in this paper utilized historical data from 2000 and 2001 (September 9 and earlier), incorporating 338 days. Thirty-eight other days were discarded due to incomplete weather or TFM initiatives (TFMI) data. We classified 62 of the 338 included days as days when the weather had a significant impact in the PA region. A significant impact was defined as a condition in which convective weather cells overlapped at least three major routes in the airspace. Out of 338 days 90 had GDP in support of SWAP or for en route weather. Not all of these GDPs were exclusively for weather in PA. Some were due to weather in other portions of the U.S., while some were for combinations of factors, such as winds followed by en route thunderstorms.

Table 1. CCFP forecast levels

Level	Coverage	Probability
Low	25 - 49%	0 – 39%
Medium	50 - 74%	40 – 69%
High	75 - 100%	70 – 100%

2.2 O'Hare Airport (ORD) Events

To contrast en route weather events specific to the northeast U.S., the ORD scenario was chosen and analyzed. Like the PA en route events, ORD weather events have high economic and operational importance for the NAS. One major distinction in the terminal area is that non-convective weather, especially ceilings and winds, is very important. Another major concern for airports, in addition to weather directly over the terminal, is convective weather over the arrival and departure fixes. When weather affects or is forecast for the terminal, the FAA may respond strategically with a GDP into ORD, or tactically with departure delays or a GS. Ceilings, winds, and severe weather over the terminal and arrival and departure fixes are the most common reasons for implementing GDPs or GSs. Typically, en route weather beyond the arrival and departure fixes is not as crucial. But on rare occasions, en route convective weather, even hundreds of miles away, can be cause for a GDP or GS into ORD.

Historical data for the period between April 1, 2000 and June 30, 2000 was analyzed. Out of this 91-day period, 38 days were classified as having a weather impact at ORD, and 24 of these days had a GDP. Non-convective weather or any type of convective weather within the terminal area defines weather impact. Convective weather impacts were also analyzed in the en route region. For this analysis, the en route region is defined as adjacent to the terminal area and extending a few hundred miles to the north, south, east and west. Weather forecast data also comes from the CCFP, and actual weather observations were taken from the National Convective Weather Forecast (NCWF) product.

2.2.1 Overall Characteristics of the ORD Events

To gain a more complete understanding of ORD weather events in general, the 38 impacted days were further categorized by means of visual classification using the Real-Time Verification System (RTVS) and information obtained from messages within the Traffic Advisory Report produced by the ATCSCC. The RTVS displays CCFP polygons along with real weather data from the National Convective Weather Detection (NCWD) product on a U.S. map. The density of activity and its duration determined the categories of low, medium, or high convective weather. Non-convective days were derived from information in the traffic advisories. It should be noted that the categories of weather are for the terminal area only, defined as the region within the four ORD cornerpost fixes.

Although some of the data sets are very small, there is a persistent increasing trend in average values for delays, diversions, cancellations, and holding as the weather becomes more severe (see Figures 1-5; weather severity increases to the right in these figures). But the data shows wide variation within each weather category. This analysis does not account for the operational TFM decisions made across different events, nor does it account for uncertainty in weather forecasts when decisions had to be made by the FAA and airlines. The next section uses a decision analysis approach based on Bayesian networks to attempt to fill these analysis gaps.

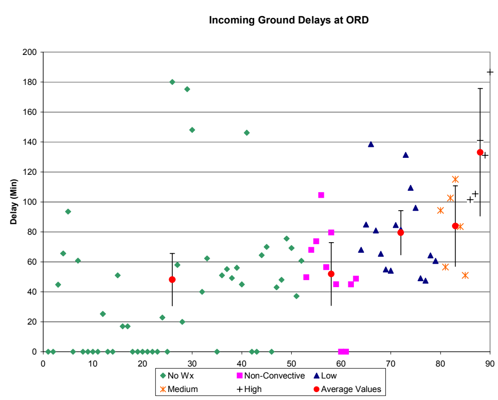

Figure 1. Average minutes of ground delay (for delayed flights) per day per weather category. Large red dots show the average value within each weather category. Data comes from OPSNET total ground delays (EDCT + GS).

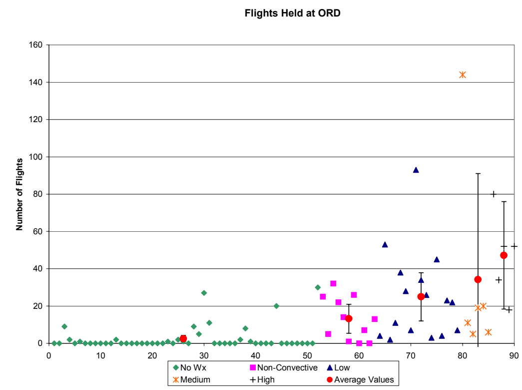

Figure 2. Number of flights held per day per weather category. Data comes from OPSNET arrival delays.

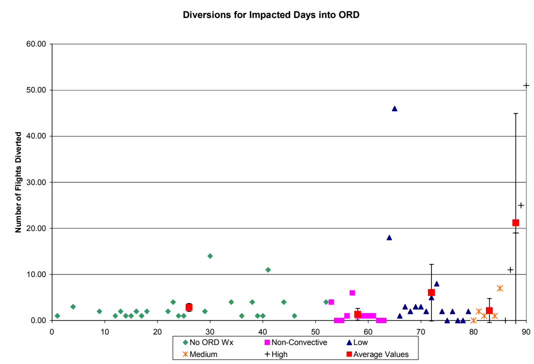

Figure 3. Number of diversions per day for each weather category. Data is derived from ETMS.

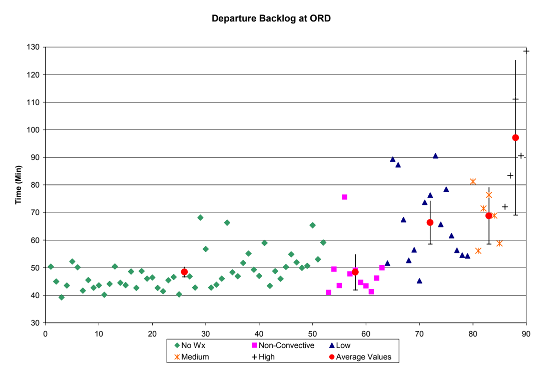

Figure 4. Average minutes of departure backlog (for delayed flights) per day per weather category. Data comes from ASPM departure delays.

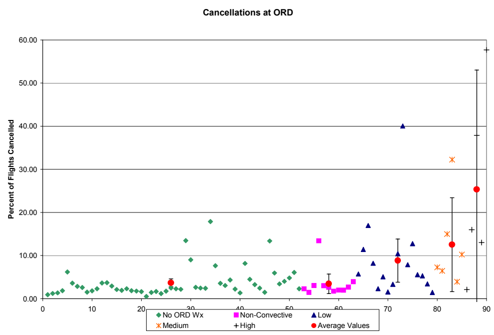

Figure 5. Percentage of flights cancelled each day per weather category. Data comes from ASPM cancellation information.

3.0 Bayesian Network Approach

3.1 Decision Making for En Route Events on a Bayesian Network

Application of decision analysis to TFM events [4] requires a representation of the decision process. The decision process for PA en route events was modeled around the Strategic Planning Telcons (SPT) collaborative process which involves the ATCSCC, certain FAA ARTCCs, Terminal Radar Approach Controls (TRACONs) and airport towers, and participating airlines. Constraints in the NAS are presented and solutions are discussed as part of the SPT process. The SPT is held every 2 hours at 1115, 1315, etc., zulu (Z) time and the FAA issues a Strategic Plan of Operations (SPO) on the following hour. The SPO includes TFMI such as GDPs, playbook reroutes, potential GSs, miles-in-trail (MIT) restrictions, and tactical reroutes.

Departure backlogs at the northeast airports are a key operational factor in these events, so this was chosen as the primary system performance factor in this analysis of weather information uncertainty. We also experimented with other possible performance factors, including departure delays for flights destined to the northeast airports, airborne holding times, diversions, cancellations, and a roll-up of various factors into an overall cost function. Results for all these factors were qualitatively similar.

The decision process was evaluated using a Bayesian network (BN) [5]. The BN encodes the probability relationships between variables on a causal network. With fairly extensive data available from TFM events in recent years, and a well-defined TFM process, we thought it might be possible to use a BN to quantify the relative effectiveness of different TFM decisions in PA en route events.

Variables in the BN include actual weather (based on NCWF detection), forecasts (based on CCFP), and TFMI. To construct our BN, we made each of our variables discrete. The following list shows each variable used in the BN, followed by the discrete values the variable can take on, along with a definition of each discrete value:

- Departure Backlog
  - 90 Plus: at least 10 flights with off-out time differences of 90 minutes or more at Newark (EWR) and Philadelphia (PHL) airports
  - 30 to 90: at least 10 flights with off-out time differences of 30 minutes or more at EWR and PHL
  - None: Other
  - (Note: LaGuardia (LGA) was omitted due to changes in the landing slot system between 2000 and 2001)
- Actual Weather (NCWF Detection)
  - Impact: at least 3 routes impacted (of J70/584/146/152/95/223/36/60/64/80) in the PA region
  - No Impact: Other
- Weather Forecasts (CCFP)
  - High, medium, or low coverage on at least three of the routes listed above
  - No forecast
- TFMI
  - GDP: GDP at EWR, LGA, or PHL designated as in support of SWAP or for en route weather
  - Playbook: One or more of the west-to-east transcontinental playbook reroutes

The BN was set up to represent the sequence of weather forecasts, weather reports, and operational decisions made in actual TFM operations. Netica software produced by the Norsys Software Corporation was used to implement the BN [6]. The BN is shown in Figure 6, with time running along the horizontal axis. Figure 6 was generated using Netica software. Times of each event and the discrete values each variable can take are shown in the boxes. Within each box, the probability of each discrete value is shown; the user can select a value with certainty, and

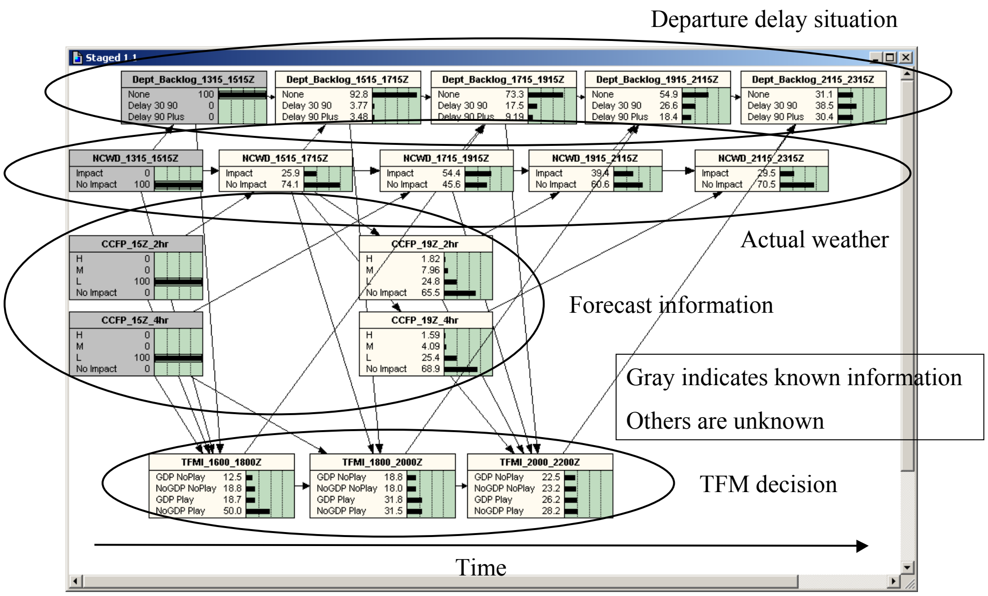

Figure 6. Diagram showing the BN for PA en route events.

the Netica software propagates the effect through the network.

Data corresponding to all 338 days of en route PA events were loaded into the BN, using the default method in Netica to update initial uniform probabilities based on new data. With these data loaded, the BN gave a picture of the estimated probability distribution of departure backlogs as a function of previous departure backlogs, actual weather, weather forecasts, and TFM decisions. Where no cases were observed, the network used the default values based on the initial uniform distributions across the possible discrete values of the variables in the model.

3.1.1 Results and Discussion

The BN captured certain limited aspects of decision making under uncertainty, but could not distinguish strategic TFM decisions in terms of their effect on system performance. The reason is twofold. First, for many important data categories, there are few or no data points based on past events. Second, there is great variation in the metrics used to assess TFMI effectiveness among events with similar traffic and weather characteristics.

For example, the distribution of actual weather in 1515 to 1715Z is conditioned on the 1500Z 2-hour forecast and actual weather in 1315 to 1515Z. There are only 2 days with medium coverage forecasts (April 20, 2000 and September 4, 2001). One day had no impact in the 1315 to 1515Z time period; and one day had an impact. This is not sufficient to determine the distribution of impact in the 1515 to 1715Z time period. This example can be overcome to an extent by using the forecast's probability and coverage to determine the probability of impact, but the forecast would have to be extrapolated to our weather classification.

The data set size of 338 days requires simplification of many relationships in the BN. For example, departure backlog is actually dependent on all TFMI implemented in prior periods. Thus, the distribution of departure backlog in the 2115 to 2315Z time period would have: (53 reasonable sets of TFMI decisions)*(3 prior departure backlog levels)*(2 actual weather levels) = 318 combinations of the conditional variables, which guarantees very small numbers of data points in many categories, given 338 days in the entire data set.

We aggregated categories in an attempt to generate larger numbers of data points in each category. For example, the BN in Figure 6 uses only the coverage element of the CCFP. Probability is also a major element of the forecast, with high, medium, and low levels. To cite another example, the TFMI variable ignores details of the decisions that are important, such as the amount of lead-time (time between the issuance of the GDP and the time arrivals are impacted). The TFMI also treats two decisions in a time period as equivalent even though they could be almost 2 hours apart.

We experimented with other BNs that aggregated variables to a much greater extent than shown in Figure 6. This increased the number of data points for some variable types, but did not increase the clarity of causal relationships in the BN. More highly aggregated BNs could not distinguish strategic TFMI in terms of effect on system performance.

Regarding the large observed variation in metrics used to assess TMI effectiveness, the first three columns of Table 2 describe types of large-scale NAS uncertainty in general. Specific examples include:

- The complicated and sensitive nature of weather's impact on traffic.
- Impacts on different combinations of routes can have different implications for traffic due to the complexity of the traffic flows. These differences can lead to variations in performance in a system near capacity that is sensitive to small changes in conditions.
- Execution of TMIs may be complicated by multiple, simultaneous actions, such as MIT imposed on top of a GDP. In addition, there are "shaky hand" effects, such as non-compliance to initiatives or ineffective communication of the start, end, or modification of an initiative.

All of these factors reduce predictability in performance and make it difficult for our BN to distinguish between the effectiveness of different TFMI. Figure 7 illustrates this issue in our particular BN.

Figure 7 shows that there is not a very clear relationship between impacts in the PA region, measured by the number of observed periods of impact as defined in the BN, and the average out-to-off delay for PHL or EWR, even conditioned on the initiatives in place. Also, we cannot see a clear relationship between performance and TFMI, even normalized for weather impact. This demonstrates a fundamental limitation in a BN decision model based exclusively on data from past events. If we cannot distinguish performance based on TFMI or weather, there is no hope of offering decision support.

However, the data used to populate the BN was adequate to show meaningful relationships between weather forecasts and actual weather, and the dependence of this relationship on weather at the time of forecast, which could be a useful input to operational decision making.

Table 2. Taxonomy of types of large-scale unpredictability in NAS operations

Type of unpredictability	Short general description	Typical NAS manifestations	Modeling approaches
Catastrophic events	Relatively rare, major events typically related to factors external to system operations	Terrorist attacks (e.g., 9/11), or labor strikes that disrupt operations on a large scale	Model representative scenarios at a high level
Complicatedness	Many factors interact and affect system behavior, making it difficult to predict what will happen	Interacting traffic flows in congested airspace and around congested airports; network effects	Model multiple interacting system elements
Sensitivity to small changes ("criticality")	System elements are often "near the edge," so behavior is sensitive to small perturbations	Demand can be near or above capacity for some NAS resources, making the system sensitive to small changes	Sensitivity analysis to changes in demand and capacity in the critical regime, where demand is near capacity
Distributed, adaptive decision-making	Multiple decision-makers interact and adapt in their own self-interest, making the system outcome hard to predict	Airspace users acting in self-interest adapt to changing circumstances, but may over-congest system resources	Agent-based or game-theoretic modeling
"Shaky-hand" effects	When actions are taken, there are large errors (accidental or deliberate) in execution	Large variance in actual aircraft arrival times compared to scheduled or GDP times; en route spacing almost independent of MIT restrictions	Agent-based or other system performance models that permit sensitivity analysis on execution of decisions
"Blurred vision" effects	Decisions are made based on imperfect information	Airline priorities are not well known to FAA; weather and demand forecasts are imperfect	Decision analysis techniques (e.g., Bayesian networks) and through sensitivity analysis on information

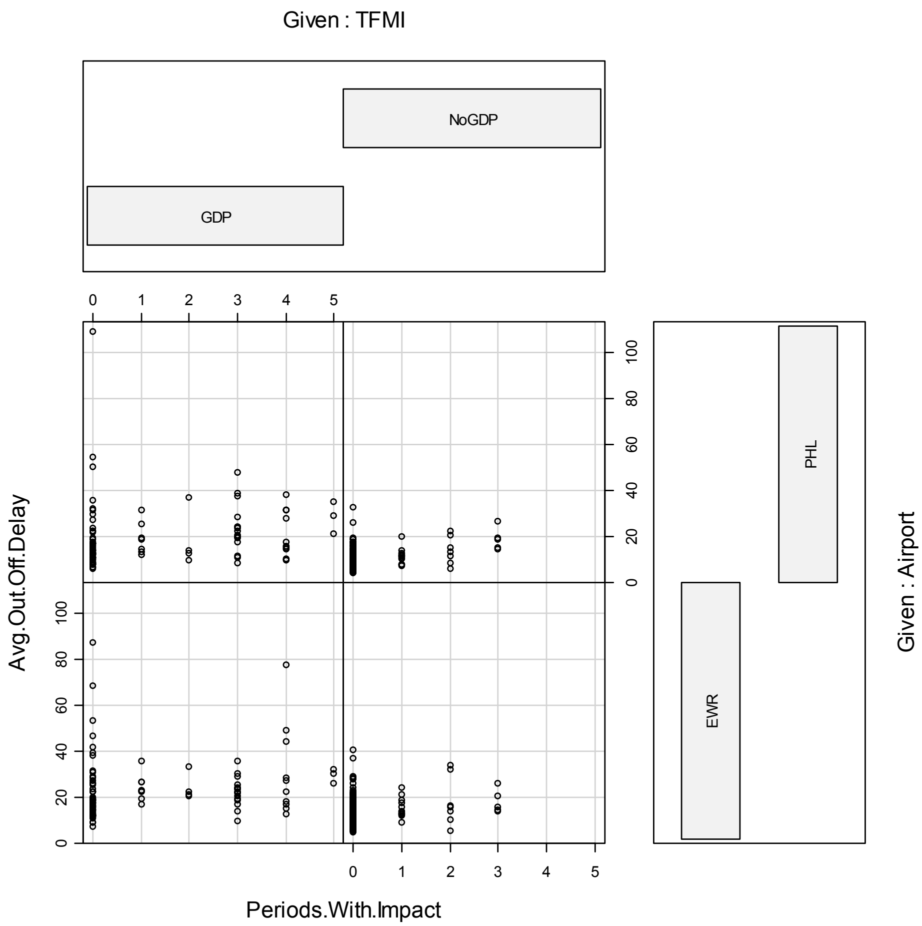

Figure 7. The chart graphs, for en route PA events, average out versus off delay during the 1315Z to 2315Z timeframe versus the number of periods from the BN with actual weather impact. The charts are conditioned on the airport and whether GDP was implemented.

3.2 Decision Making for ORD Airport Events on a Bayesian Network

In order to create a BN for ORD with reasonably populated events, a number of simplifications were made. First, the TFM decision was limited to either implementing a GDP or waiting; thus, decisions to execute a playbook or GS were not considered. Second, no distinction was made between probabilities of occurrence and coverage levels in weather forecasts, i.e., there was either a forecasted weather impact or no impact. Actual weather events were also categorized as having an impact or no impact. The airline decision, in response to any imposed GDP, was to make either few or many cancellations. These decisions and events were analyzed for the period of April 1, 2000 through June 30, 2000.

The performance metrics used to illustrate the impacts of weather and decisions include en route holding, incoming ground delays, diversions, departure backlogs, and total delays. They were taken from operations network (OPSNET), Aviation System Performance Metrics (ASPM), and enhanced traffic management system (ETMS) data; each day was then classified as having either high holding or low holding, high ground delays or low ground delays, etc. Specifically, high en route holding, taken from OPSNET arrival delays, was defined as any day in which more than 25 flights had any holding into ORD. High incoming ground delays, from OPSNET total ground delays (a combination of estimated departure clearance time (EDCT) and GS delays), are days with average delays greater than 60 minutes. Departure backlog data was taken from OPSNET departure delays. A high departure backlog day was defined to be a day in which more than 75 flights had any kind of departure delay. High total delays, from OPSNET total delay times, had an average of more than 60 minutes of delay. Diversions were derived from ETMS data and days with more than 10 diverted flights were categorized as high. Information on cancellation decisions was obtained from ASPM data. Days with high cancellations were defined as those in which more than 10% of arriving flights were cancelled (the percentage was calculated as cancelled arrivals/scheduled arrivals for metric computation).

As for the PA en route event BN, the ORD terminal event BN does not yield useful results for strategic decision making. The BN indicates that the best decision is to always make few cancellations and not implement a GDP regardless of weather forecasts or actual weather impacts. The input data for some of the nodes was also surprising. For instance, there is a greater chance of having high airborne holding when there is weather, many cancellations, and high incoming ground delay than when there is weather, low incoming ground delay, and few cancellations. One would expect that when there are many cancellations and flights held on the ground there would be less need for airborne holding. Similarly, when given a GDP with many cancellations there is a greater chance of having high delays than when given a GDP and few cancellations. One explanation is that *most* severe weather days fall into the high delay/many cancellations or GDP/many cancellations categories, and thus *all* metrics are high as a result of the weather severity rather than as a result of any TFM decision. If there had been no GDP or few cancellations on the most severe weather days, holding and other metrics would probably be even higher. However, there was only one such day with severe convective weather, few cancellations, and no GDP. This lack of data for specific event types is similar to that observed for PA en route events.

3.3 A More Focused Analysis of PA Events

3.3.1 Approach

This analysis isolated all days in Spring/Summer 2002 that had weather in the PA en route region and where GDP SWAP was implemented. On these days, the beginning of the estimated arrival period was identified. A weather index for that time was computed. The weather index tracks weather impacts at specific locations in the NAS. The locations were obtained from the SPT/Severe Weather and Route Management document, available on the FAA's Operational Information System (OIS). In total, 308 navigation aids (NAVAIDS), 231 fixes, 28 airports, and 78 jet routes (1107 route segments) were tracked. Scores for regions were computed by counting the number of locations impacted in the region at a given time. The PA region selected for analysis overlaps the border between the New York and Cleveland ARTCCs (ZNY and ZOB).

The analysis attempts to isolate the effectiveness of the plan by evaluating how well the plan was executed without significant modifications. To do this, each GDP was evaluated manually. The frequency, scope, and reasoning behind each GS during that program were considered. Each GDP SWAP event was classified as low, medium, or high, indicating the extent of GS overlap in the GDP. GS for en route weather during a GDP were used to compute the overlap. However, GS implemented at the issuance of a GDP were not included in the overlap since they may be part of the strategic plan to help set up the GDP. Also, GS for terminal thunderstorms during a GDP for en route weather are not included, since GS are often expected under those conditions.

3.3.2 Discussion

This more focused analysis has several advantages over the comprehensive BN approach described earlier. First, no CCFP data are included, minimizing the data requirements. Second, GDP categorization into 2-hour bins is no longer an issue. The GDP timing is evaluated relative to the weather. Third, the weather is treated as a continuous variable. This does not account for all of the complex issues, but it does significantly improve the distinction between weather events of different scopes.

The main disadvantage is that the analysis no longer integrates the forecast into the analysis. This approach leaves the forecast integration to a second step where forecast uncertainty is incorporated into the decision process. A specialist faced with a GDP SWAP decision must make an estimate of his/her belief in the forecast and balance the perceived benefit of the GDP SWAP against the possible cost of a GDP SWAP if the weather does not materialize as forecast. Unfortunately, we cannot provide the performance benefit for the GDP SWAP or the wait-and-see alternative, but our analysis can support TFM specialists' expectations about how the GDP will evolve. We also cannot give weather index scores for the CCFP based on the weather forecast since the forecast is not fully defined in terms of probabilities. For example, a medium probability forecast of medium coverage leaves open the possibility of low coverage or no coverage at all. These probabilities are not explicitly defined in the forecast. Also, the probability and coverage ranges in the CCFP are enormous. The low probability designation covers 1% to 39% and the low coverage designation covers 25% to 49%, which includes a wide range of different situations.

Finally, the methodology for scoring the execution of a GDP SWAP event needs to be formalized. Specific rules on acceptable and unacceptable GS within a GDP should be developed.

3.3.3 Preliminary Results

Several days were removed if the GDPs were put in at widely varying times, often due to other conditions at the terminals, such as low ceilings or wind. The results are presented in Table 3.

Table 3. GDP/GS overlap analysis results

Date	GDP/GS Overlap	Score at GDP Start
5/9/02	Low	0
4/28/02	Low	10
5/28/02	Medium	8
6/26/02	Medium	11
7/19/02	Medium	18
6/27/02	High	28
7/23/02	High	26

There are too few points to draw any solid conclusions from the data, but the results do offer some insight into effective implementation of GDP SWAP. Higher weather index scores at the time of GDP implementation correlate with increasing GDP SWAP execution problems.

3.4 Approaches to Augmenting Data from Actual Events

The BN approach, while appealing in its completeness and simplicity, requires far too much data to be practical. The inability to account for all relevant factors in the system makes TFM performance assessment an elusive target. Two possibilities to address this problem are to incorporate subjective input to augment the data, and to augment the data with modeling and simulation results.

TFM shows wide variability in outcome metrics across different events. Generic approaches to modeling complex systems with such "fat-tailed" outcome distributions include models of self-organized criticality (SOC) [7], Highly Optimized Tolerance (HOT) [8], and the NK model of biological evolution [9]. However, these models would need considerable adaptation to be applicable to TFM modeling. We propose modeling TFM events in reference to the taxonomy of sources of unpredictability in the NAS. Table 2 lists an approach to modeling each different source of unpredictability; these would need to be combined to form a complete picture of predictability and uncertainty in TFM events.

4.0 Conclusions

System-level performance metrics are highly variable across individual TFM events, but there are recognizable patterns. Data from past TFM events is not sufficient to distinguish between strategic TFM decisions in a Bayesian decision network, in terms of metrics based on overall delays, cancellations, diversions, and departure backlogs. However, our results show that useful information can be extracted from data on past TFM events by focusing on specific elements of the strategic TFM process rather than the entire process comprehensively.

The difficulty in creating a usable Bayesian decision network highlights how difficult it is to learn to make better strategic TFM decisions from past decision-making experience. At a tactical level the TFM system is remarkably adaptive in responding to changing circumstances. This implies that models of TFM decision making should emphasize tactical adaptation and learning with respect to specific elements of the strategic TFM process, rather than learning at a comprehensive strategic level.

Further research is needed to complete the link between the strategic and tactical levels of TFM. Modeling and simulation may provide a useful framework to do this, and we present a taxonomy of sources of uncertainty and modeling approaches to address these (Table 2). Use of relatively simple models may be the best way to proceed in this research [10]. The issues addressed here are pervasive in complex adaptive systems and may be of central importance to improving aviation operations to meet the future needs of the flying public.

Acknowledgments

The work described in this paper was funded as MITRE-sponsored research and development. We gratefully acknowledge the help and suggestions of Dr. Brad Hargroves, Dr. Craig Wanke, Mark Huberdeau, Jack Brennan, and Joseph Hollenberg, all of MITRE CAASD. We also acknowledge help and encouragement from the FAA ATCSCC.

NOTE: The contents of this material reflect the views of the authors and/or the Director of the Center for Advanced Aviation System Development. Neither the FAA nor the Department of Transportation makes any warranty or guarantee, or promise, expressed or implied, concerning the content or accuracy of the views expressed herein.

References

[1] Ball, M. O., R. L. Hoffman, D. Knorr, J. Wetherly, and M. Wambsganns, 2000, "Assessing the Benefits of Collaborative Decision Making in Air Traffic Management," *3rd USA/Europe Air Traffic Management R&D Seminar*, Napoli, Italy.

[2] Ball, M. O., T. Vossen, R. Hoffman, 2001, "Analysis of Demand Uncertainty Effects in Ground Delay Programs," *4th USA/Europe Air Traffic Management R&D Seminar*, Santa Fe, New Mexico, U.S.

[3] Campbell, K. C., W. W. Cooper, D. P. Greenbaum, and L. A. Wojcik, 2001, "Modeling Distributed Human Decision Making in Air Traffic Flow Management Operations," in *Air Transportation System Engineering*, ed. G. L. Donohue and A. G. Zellweger, Reston VA, U.S., AIAA, pp. 227-237.

[4] Wojcik, L. A., 2001, "Three Principles of Decision-Making Interactions in Traffic Flow Management Operations," *4th USA/Europe Air Traffic Management R&D Seminar*, Santa Fe, New Mexico, U.S.

[5] Heckerman, D., 1996, "A Tutorial on Learning with Bayesian Networks," Microsoft Research Technical Report MSR-TR-95-06, Redmond, Washington, U.S., March 1995 (revised November 1996).

[6] Norsys Software Corp., 1997, "Netica Application for Belief Networks and Influence Diagrams, User's Guide," Version 1.05 for Windows, Vancouver BC, Canada, Norsys Software Corp.

[7] Bak, P., 1996, *How Nature Works: the Science of Self-Organized Criticality*, New York, Springer-Verlag.

[8] Carlson, J. M. and J. Doyle, 1999, "Highly Optimized Tolerance: A Mechanism for Power Laws in Designed Systems," *Physical Review E 60*, pp. 1412-1427.

[9] Kauffman, S., 1993, *The Origins of Order: Self-Organization, and Selection in Evolution*, New York, Oxford University Press.

[10] Wojcik, L. A., 2001, "Models to Understand Airline and Air Traffic Management Authority Decision-Making Interactions in Schedule Disruptions: From Simple Games to Agent-Based Models," in *Handbook of Airline Strategy*, ed. G. F. Butler and M. R. Keller, New York NY, U.S., McGraw-Hill, pp. 549-575.

Key Words

Traffic flow management, decision analysis, Bayesian networks.

Biographies

Joshua W. Pepper is Senior Staff in MITRE CAASD. He has worked at MITRE for 2 years, providing data support for Quality Assurance at the ATCSCC. He holds an M.A. in Applied Mathematics (Operations Research) from the University of Maryland, College Park.

Kristine R. Mills is a Senior Simulation and Modeling Engineer in MITRE CAASD. She has worked at MITRE for 2 years, providing analysis and algorithm development for various Terminal RNAV projects. She holds Ph.D. and M.S. degrees in Applied Mathematics from Northwestern University.

Leonard A. Wojcik is Project Team Manager in MITRE CAASD. He has worked at MITRE for over 20 years. His group applies system-level air traffic simulation to key aviation investment and operational questions. He holds a Ph.D. from Carnegie-Mellon University in Engineering and Public Policy.