

TRAFFIC FLOW MANAGEMENT MODELING AND OPERATIONAL COMPLEXITY

Brendan Hogan
Leonard A. Wojcik

Center for Advanced Aviation System Development (CAASD)
The MITRE Corporation
7515 Colshire Drive
McLean, VA 22102 U.S.A.

ABSTRACT

Traffic Flow Management (TFM) actions are commonly used to mitigate capacity/demand imbalances within the National Airspace System (NAS). Modeling TFM events has proven challenging in the past, partly because of weather forecast uncertainty, and partly because of the complexity and unpredictability associated with highly-interrelated traffic patterns and distributed decision-making in the NAS. In this paper, we present results of a simulation of a NAS TFM event in which weather effects are relatively small. This facilitates interpretation of the similarities and differences between simulation results and the actual event in terms of NAS operations and decision making, with relatively small weather-related complications. We conclude that TFM modeling shows promise as a tool to aid post-event TFM analysis, but the complex operational factors impose limits on the predictability of outcomes in TFM events. A CAASD-developed fast-time network simulation of the NAS was used for this analysis.

1 INTRODUCTION

The National Airspace System (NAS) is an extremely complex network that handles over 60,000 flights each day. When demand for a resource in the system is expected to exceed the capacity, Traffic Flow Management (TFM) actions are taken to resolve the problem. These TFM actions can range from Miles-In-Trail (MIT) restrictions for en route flights, to Ground Delay Programs (GDPs) and Ground Stops (GSs) for flights that have not yet departed. MIT restrictions enforce a minimum longitudinal separation between aircraft that exceeds the minimum separation standard. A GDP adjusts departure times for aircraft scheduled to arrive at the affected airport, in order to reduce the arrival rate. A GS simply holds flights on the ground that are due to arrive at an airport, for some time period. Each of these actions has different

delay implications for the airspace users, as well as different effects on the demand imbalance it is meant to address. Some examples of situations that cause a need for TFM actions are when weather conditions or runway outages limit the capacity of an airport. Modeling TFM events has proven challenging in the past, in part because of the uncertainty associated with weather forecasts, and in part because of the complexity and unpredictability associated with highly-interrelated traffic patterns and distributed decision-making in the NAS (Campbell et al. 2001). Uncertainty in weather forecasts can be modeled to some degree with such methods as decision analysis, but the NAS complexities are more difficult to capture (Pepper, Mills, and Wojcik 2003).

In this paper, we present the results of a simulation of a TFM event that occurred at Newark airport on August 23, 2003. This event is of particular interest because the TFM actions were motivated by a planned runway outage at the airport and the weather at the airport was predictably good throughout the day. Thus, comparison between modeled and actual outcomes can be used to characterize the complexity of NAS operations and decision making, with relatively limited complications arising from weather uncertainty. Our approach throughout this work is a parallel effort between modeling the TFM event and understanding how the details of the actual event unfolded on the analysis day. We believe this is the best way to answer questions of what we can and cannot usefully model with respect to TFM actions.

2 MODELING ASSUMPTIONS

The questions that this simulation study aimed to address are those that pertain to airport-specific demand/capacity imbalances at Newark International Airport (EWR) and the TFM actions taken to handle them. In actual operations, the behavior of flights is affected by airspace issues and congestion in other areas of the NAS aside from EWR, but it is very challenging to model these effects in a useful

way. Based on our operational and past modeling experience, we decided that it would not be worthwhile to begin the analysis by attempting to include these effects in the model. Rather, we kept the model as simple as possible consistent with capturing the essential features of the problem by modeling only the flights going into or out of EWR on Saturday, August 23, 2003.

In an effort to show EWR-related delay propagation effects through the day, flight itineraries are linked whenever possible based on the tail number of the aircraft used for the flight. Our source of this information is the Department of Transportation's Airline Service Quality Performance (ASQP) database. ASQP covers flights within the continental United States on airlines having at least one percent of the total scheduled domestic passenger revenues (Office of Airline Information 2004). At a large international airport such as EWR the ASQP data covers only about 65 percent of the flights. For those flights that are not covered by ASQP we are not able to realistically link them into multi-leg itineraries. Nonetheless, we included them in the simulation as unlinked (single-leg) itineraries to preserve the traffic volume levels at the airport. Significant categories of traffic that are missing from ASQP are international, cargo and general aviation flights, as well as small commercial carriers. These remaining flights for our scenario were taken from the Enhanced Traffic Management System (ETMS) database (Volpe Center 2000). The ETMS data available at The MITRE Corporation's Center for Advanced Aviation System Development (CAASD) is the Aircraft Situation Display to Industry (ASDI) data feed, which includes all non-military flights operated under Instrument Flight Rules (IFR). This database of IFR traffic is in effect the entire set of flights that the Federal Aviation Administration (FAA) air traffic controllers must handle. The ASQP and ETMS data sets were merged to form a complete set of flights.

As mentioned previously this analysis day was chosen in part because of the relative lack of complicating weather across the NAS. For that reason the airport capacities were modeled as the arrival and departure rates for Visual Meteorological Conditions (VMC) as determined by an FAA/CAASD capacity benchmarking study (FAA 2001). During the portion of the analysis day that had the runway closure at EWR, the reduced capacity of the airport strongly affects any arrival or departure delays that may result. Logically, any delay-related simulation output is very much a function of the airport capacity assumptions that are made on the input side. For this analysis we modeled the runway closure with a reduction in the arrival and departure capacity at EWR. In the actual event, there were differing opinions amongst the stakeholders as to what the actual capacity was during this outage. These opinions, how they varied dynamically as the event unfolded, as well as our modeled reduced capacities are discussed in detail in Section 3. We mention this here to emphasize that even on a relatively

clean weather day with a planned capacity reduction there are complex issues associated with assessing an airport's ability to handle traffic.

3 OPERATIONAL COMPLEXITY

Modeling a TFM event is constrained by the ability of the modelers to understand and appropriately capture the operational complexities associated with that event. There are often several stakeholders affected by the TFM situation, each of whom likely has unique objectives and opinions on what the best course of action may be. For this example scenario of the runway outage at EWR, the main players involved include the New York Air Route Traffic Control Center (ZNY ARTCC), the New York Terminal Radar Approach Control (N90 TRACON), the Air Traffic Control System Command Center (ATCSCC), which does national-level TFM, the EWR control tower, as well as Continental Airlines (COA), the dominant carrier at EWR. (ARTCCs and TRACONs are FAA air traffic management (ATM) facilities.) In addition, other parties that may be affected by the situation include airlines such as Delta and United, and surrounding ARTCCs such as Washington (ZDC) and Boston (ZBW).

Decisions regarding TFM actions are made through a series of planning teleconferences among the stakeholders that take place several times a day. It is in these meetings that different plans are discussed and often conflicting opinions are heard between the players. It is common for an airline to be more aggressive than FAA facilities in what they feel the arrival capacity of an airport should be, as well as be more likely to assume the risk of airborne holding associated with an aggressive strategy. The airport control tower and the TRACON controllers are frequently more conservative in their capacity estimates and planning of actions to handle demand imbalances. All of this is coupled with uncertainty in the predictions of both demand and capacity. On the demand side, the stochastic nature of the departure and en route delays of flights destined to the focus airport contributes to the planning complexity. On the arrival side, most of the uncertainties arise from weather effects.

The baseline day for this study was chosen in part because of the planned capacity reduction at EWR and in part because of the good weather across the NAS. To put this in perspective, August 23, 2003 had just 156 weather delays in the NAS, compared to the daily average of 855 weather-related delays for all of June – August 2003, according to data from the FAA Operations Network (OPSNET) database (FAA 2004). Despite this, the weather at EWR was an issue in the decision making of controllers and ultimately the TFM actions that were taken. There were varying estimates throughout the day as to what the arrival capacity at EWR would be during the runway closure that was to start at 13:30 local time and

continue for the rest of the day. As our simulation results show, the difference in delay effects would be significant based on whether the airport was at the upper or lower bounds of their estimated arrival capacity during the time of the runway outage. We believe that these varying capacity estimates were due to the changing wind conditions at the airport during the day.

The preferred airport configuration for handling the runway outage was to use runway 4R for departures and runway 29 for arrivals as shown in Figure 1. Using runway 29 for arrivals implies a heading on final approach of roughly 290 degrees relative to north. If there is a strong wind out of the north (360 degrees), this would represent a prohibitive 70-degree crosswind for flights landing on runway 29. This situation would cause the airport to settle for mixed-operations on the single practical runway 4R, and would decrease the capacity accordingly.

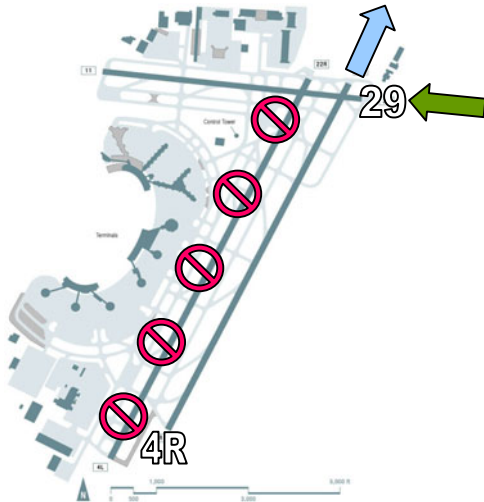


Figure 1: EWR Preferred Configuration during Runway Outage

Figure 2 shows the hourly wind vectors on the analysis day at EWR and the Official Airline Guide (OAG) arrival demand, along with the predicted capacity during the time of the runway outage. The wind data was taken from the Aviation System Performance Metrics (ASPM) database provided by FAA/APO (FAA, 2004), and the predicted arrival capacities were taken from ATCSCC logs. As the OAG schedule represents unconstrained demand, when compared with the capacity of the airport, it is clear that the 14:00, 16:00, and 18:00 hours are the reasons a GDP was required on this day. Note that in the early afternoon, the winds shift from North to North-West, a more favorable situation for utilizing both runways 29 and 4R. It is about this time that the airport changes to a more optimistic predicted arrival capacity of 38 from 34. This increased rate was implemented in a revised GDP for EWR, and might ultimately have been the cause of airborne holding delay later in the day. There is evidence

in the ATCSCC logs from the day that in hindsight some controllers felt the GDP 38 rate “might have been a little aggressive since we had to hold ZDC off and on.”

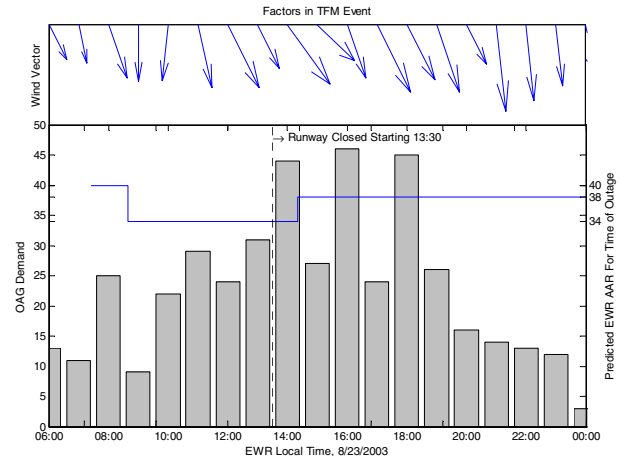


Figure 2: Wind Driven Adjustments to Capacity Prediction

4 RESULTS

4.1 Baseline Validation

As the first step in this modeling effort, we assessed our ability to produce simulation results that approximate performance data from the actual TFM event on August 23, 2003. Since the focus of this study is on the TFM issues of the demand/capacity imbalance at EWR, we used airport-specific metrics for this validation. Figure 3 shows data from the actual event in a CAASD-developed airport-specific display format (Campbell, Pepper, and Yankey 2002). At this point, we will thoroughly explain the sections of this figure as we use this format several times in this paper. The top chart in Figure 3 shows the hours of any GPs or GDPs that took place at the analysis airport. In this case there was just one GDP for EWR, from 14:30 until 20:00 local time. In the next and largest chart OAG scheduled (connected points) and actual (bars) operations are both represented. The convention in this and following figures of the same format is that arrivals are shown on the top half of the chart and departures on the bottom. The units on these plots are flights per 15-minute time bin. Also shown in the purple line is the airport’s called arrival rate as reported in the FAA’s Aviation System Performance Metrics (ASPM) database. Note that the called arrival capacity drops from 40 flights per hour to 34 at 13:30 Eastern Standard Time (EST) corresponding to the time that the runway closed. The wavy nature of the capacity line beginning at 13:30 is not an actual variation but an artifact of displaying 15-minute bins for a rate not evenly divisible by four. The third chart from the top displays diversions as well as departure and arrival cancellations at the analysis airport.

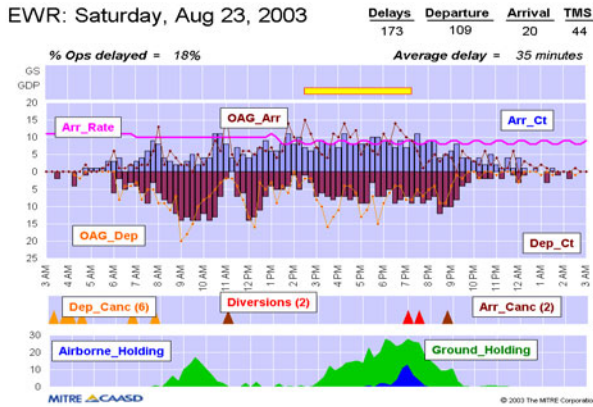


Figure 3: Data from the Actual Event

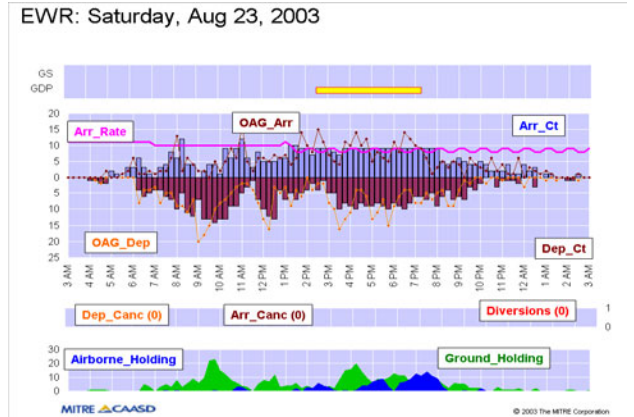


Figure 4: Data from the Baseline Simulation

The bottom chart in Figure 3 shows ground and airborne holding at EWR. There is an important distinction to be made between flights subject to a GDP, and flights that experience “Ground Holding” as represented in this chart. The flights shown here as being in Ground Holding are EWR departures that have pushed back from the gate but have not yet taken off. In contrast, flights that are part of a EWR Ground Delay Program are Newark arrivals deliberately held at their origin airport to meet a TFM objective. The airborne holding displayed in Figure 3 represents EWR arrivals experiencing airborne queuing delay. For that reason, the most important elements of the graph for our validation purposes are the actual arrival and departure counts compared to the capacity of the airport, and the airborne holding.

Note that this was a relatively clean day for the analysis airport with few diversions or cancellations. There was a large morning and early-afternoon departure push, followed by a period of more steady operations due to the afternoon runway closure and associated GDP. In the morning departure push the airport appeared to be operating at its capacity since there is a 45-minute span where the scheduled departures exceed the actual, concurrent with an accumulation of ground holding. Despite the GDP with a constant rate of 38 arrivals per hour, there was some variation in the actual arrival counts. This was due to normal stochastic effects such as time made up or lost on departure or en route. The most noticeable aspect of this figure is the afternoon peak in both airborne holding and ground holding. We will explain our efforts to understand this in more detail in the next section.

For comparison, the results of our baseline simulation are shown in the same format in Figure 4. Note that we did not model diversions or cancellations in this baseline simulation, so those fields are blank in the display. Also, the OAG scheduled departures and arrivals, in addition to the called arrival rate, are shown merely for illustration purposes as they have the same values as the actual event and we do not explicitly model them.

Note that there is a spike of airborne holding at about 19:00 local time (23:00 Greenwich Mean Time (GMT)) that occurs in both the actual event (peak 13 flights) and our simulation results (peak 13 flights). This was reassuring from a validation standpoint. The ground holding in the simulation (peak 22) mirrors that in the actual event (peak 18) during the morning departure push; however it is missing in the afternoon. A main difference between our simulated departures in the afternoon compared to those in the morning is that the early departure push is mostly comprised of the first leg of the day’s itinerary for each aircraft. In the afternoon the departures are almost all by aircraft that have flown preceding legs that day and thus can propagate delays from one airport to another. As described in Section 2, our methods for linking flight legs into aircraft itineraries are imperfect and some flights are clearly not connected when they should be. If these flights are not linked properly in the simulation, they will simply depart the airport at their originally scheduled time. This effect of the missing flight itineraries is also multiplied somewhat by our assumption in Section 2 to include only flights into and out of EWR. Aircraft itineraries that include triangle routes (A to B to C to A) or any polygon shaped route will be improperly linked by this method. Aside from these unlinked flights, the discrepancy in ground holding may have been due to a drop in departure capacity that was not modeled in the simulation. However, as we explained earlier in this section the TFM actions we seek to model have impacts on EWR arrivals, therefore we are less concerned with matching simulated to actual results in the ground holding metric defined in Figures 3 and 4.

This leads into the question of why there was airborne holding in the first place on a clear weather day in which a GDP was implemented in response to a planned capacity reduction. Also, given that the airborne holding occurred in the actual event, is that holding being appropriately captured in the model? As discussed in Section 3, there are many operational complexities in the system we are studying. In this work we aim to identify those complex-

ities which we can and cannot model effectively. To get at those questions we look a little deeper into the airborne holding that developed on the analysis day. Figure 5 shows a snapshot of flights destined to EWR at 19:00 EST, during that peak of airborne holding.

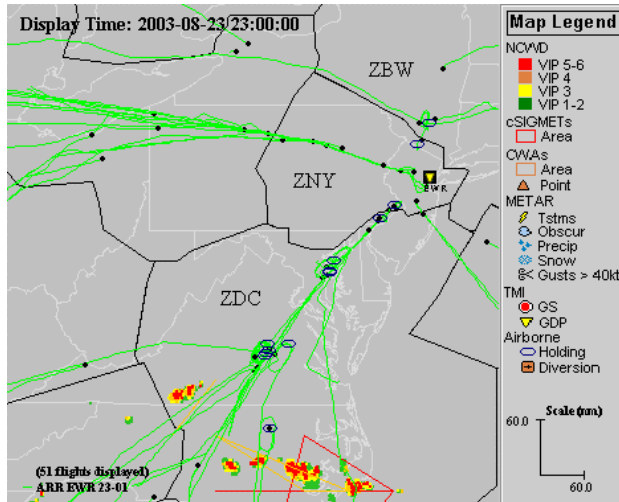


Figure 5: Snapshot of Airborne Holding among Flights Destined to EWR at 23:00 GMT (19:00 EST)

This is a CAASD-developed display that portrays multiple layers of data useful for post-event evaluation (Yankey 2003). Each black dots represents a flight destined to EWR and the green lines are the tracks those flights have taken up to that point. The yellow triangle at EWR indicates that a GDP was in place at that time. Of particular interest for this work are the blue ovals representing flights that are currently in airborne holding. This is determined with a CAASD-developed algorithm based on the geometric properties of the four dimensional flight tracks, based on radar data from ETMS (Gordon and Yankey 2002). Also shown in Figure 5 with the green, yellow, orange, and red areas is problem weather of increasing intensity from the National Convective Weather Detection (NCWD) product. Convective weather is commonly associated with thunderstorms and consists of vertical movement of the air that is extremely hazardous in aviation and, therefore, must be avoided in planning by air traffic control (ATC) and the airlines. If convective weather develops over the approach route to an airport, for example, the arrival rate would likely drop and flights could queue up in airborne holding or TFM actions such as MITs or a GDP could result. As shown in Figure 5, there is some convective weather over North Carolina in southern ZDC; however it is not blocking the routes of the EWR arrivals and does not appear in any way involved with the airborne holding that is observed.

Also note that the airborne holding is spread out from central ZDC up to southern ZBW, with little of it in ZNY itself. Initially we were still uncertain as to whether the

airborne holding that resulted was due to some inefficiencies of the GDP that was implemented or possibly due to other complicating effects such as en route congestion. After consulting with subject matter experts, we concluded that this airborne holding was in fact due to terminal capacity limitations at EWR. The reason that the holding is spread out is that the airspace in ZNY, and N90 in particular, is complex and congested leaving controllers no room to absorb delays with airborne holding. As a result, the routes into EWR through ZDC and ZBW back up and absorb the delays further upstream.

Based on this analysis, we conclude that since the airborne holding in the actual event was a result of EWR terminal constrains we have captured this holding appropriately in the simulation.

4.2 Simulation Excursions

To explore issues of a specific capacity assumption on the presence or absence of a GDP, we present the simulation results of Figures 6 and 7. Figure 6 shows output data from a simulation in which there is an assumed capacity during the runway outage of 34 arrivals and 34 departures per hour. In addition, a GDP is modeled as it was implemented in the actual event. That is, the GDP originally called for a 34 rate and then increased to a 38 rate. Our results indicate that a queue of airborne holding develops in the late afternoon, for a total of 3700 minutes of air holding among affected flights.

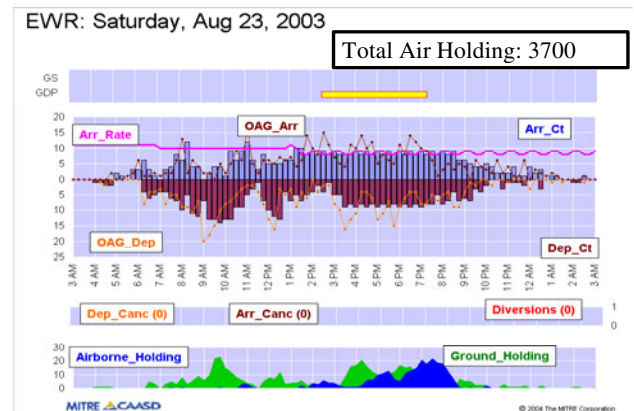


Figure 6: Simulated GDP as Run Assuming Capacity during Outage of 34 Arrivals and 34 Departures

For comparison, we present in Figure 7 the results of a simulation in which the reduced capacity is also assumed to be a fixed 34-34 rate for arrivals and departures, however no GDP is implemented. As expected, with no GDP to thin out the arrival demand a more severe queue of airborne holding develops for a total air delay of 5000 minutes among affected flights.

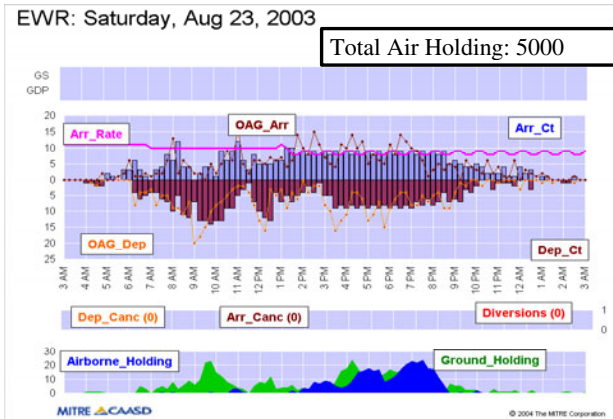


Figure 7: Simulated No GDP Assuming Capacity during Outage of 34 Arrivals and 34 Departures

4.2.1 Proposed Approach for Comparing between Scenarios

With several different TFM simulations, it would be helpful to develop a technique to quantitatively compare results between them. We propose here some variables that would likely add to the value of any such comparison. Note that these variables and any associated metrics would only be considered from the subset of flights that are related to the TFM event at hand. To use the example of the baseline day in this study, we would only compute the metrics for the subset of flights scheduled to arrive at EWR during the runway outage. These variables could include the total holding for all flights in the set, the fraction of flights with total holding greater than some threshold, the total airborne holding for all flights in the set, or the fraction of flights with airborne holding greater than some threshold. The threshold on the total holding for a flight is meant to represent the point at which delays for a flight start to become fairly painful. The threshold on the airborne holding for a flight is meant to represent the point at which long airborne delays begin to cause flights to divert to an alternate airport, and therefore greatly complicate things for the airline and passengers involved. These metrics could then be multiplied by cost scalars that enforce how relatively painful each of the conditions is for the scenario.

There are strengths and weaknesses of attempting to quantify a TFM scenario in this way. The main strength is that it is possible to roll-up the variables into a single cost function for the entire simulation. This would allow the analyst to easily compare between alternative situations (Campbell et al. 2001). The main weakness is that the costs associated with different types of delay vary greatly on a flight-by-flight basis. Therefore, it is difficult to develop a realistic cost function, or even one that appropriately captures the desired system effects.

An additional benefit of creating a cost function is that it can be applied to all flights involved with a particular

event, or it can be applied to each airline separately. If the cost is calculated for each air carrier, it can be used to perform studies and draw conclusions about the equity implications of various TFM events. Assuming that a cost function can be developed and validated to represent the interests of the stakeholders, it is possible to use that cost as an objective function to be minimized through the use of simulation optimization. If an analyst is trying to determine the ideal rate for a GDP, or whether a GS or GDP should be applied to Tier 1 (just airports in adjacent centers) or all of the NAS, reasonable outcomes would occupy a sample space and simulation optimization could find the solution associated with the minimum cost. A simulation optimization could be designed for other objectives as well. For example, it may be of interest to find the TFM action that would maximize equity among the major airlines at an airport. This could be arranged by creating an objective function from, for example, the sum of squared differences across each airline's cost.

5 CONCLUSIONS

We have presented a TFM modeling analysis in which we assess which elements of the operational complexities we are able to effectively capture in a simulation. For this work we selected an analysis day in which a TFM action was necessary due to a planned runway outage at EWR airport. This day had the added benefit to our modeling work of predictably good weather throughout the day. Even on this relatively good weather day, there was some operational uncertainty over what the capacity of the airport would be. This uncertainty likely caused some airborne holding despite the GDP that was implemented. We conclude that TFM modeling shows promise as a tool to aid post-event TFM analysis, but the complex operational factors impose limits on the predictability of outcomes in TFM events.

Future research work in this direction could include further development and validation of a cost function for comparison between alternative TFM scenarios, followed by a simulation optimization for a given TFM event based upon that cost function.

ACKNOWLEDGMENTS

The authors would like to acknowledge the invaluable consultation and modeling advice of MITRE CAASD's Bradley Hargroves, William Trigeiro, and Michael White, without whose help this work would not have been possible.

REFERENCES

- Campbell, K. C., W. W. Cooper, D. P. Greenbaum and L. A. Wojcik, 2001, "Modeling Distributed Human Decision Making in Traffic Flow Management Operations," in *Air Transportation System*

- Engineering*, ed. G. L. Donohue and A. G. Zellweger, pp. 227-237, American Institute of Aeronautics and Astronautics (AIAA), Reston, Virginia.
- Campbell, K. C., J. W. Pepper, and M. A. Yankey, 2002, "Design of Data Analysis Infrastructure," MP02W0000133, The MITRE Corporation, McLean, Virginia.
- Federal Aviation Administration, 2001, "Airport Capacity Benchmark Report 2001," [online]. Available online via <http://www.mitrecaasd.org/library/general/airport_capacity_benchmarks.pdf>, (Accessed May 28, 2004).
- Federal Aviation Administration - Aviation Policy and Plans (APO), 2004, "APO Data Systems," [online]. Available online via <<http://www.apo.data.faa.gov>>, (Accessed May 27, 2004).
- Gordon, L. and M. Yankey, 2002, "Airborne Holding Information on a Next-Day Basis," *Journal of Air Traffic Control*, Volume 44, Number 2.
- Office of Airline Information - Bureau of Transportation Statistics, 2004, "Sources of Air Carrier Aviation Data," [online]. Available online via <<http://www.bts.gov/oai/sources>>, (Accessed May 28, 2004).
- Pepper, J. W., K. R. Mills and L. A. Wojcik, 2003, "Predictability and Uncertainty in Air Traffic Flow Management," in *5th USA/Europe Air Traffic Management R&D Seminar (ATM-2003), Metrics and Performance Management*, Budapest, Hungary.
- Volpe Center, 2000, "Aircraft Situation Display to Industry: Functional Description and Interface Control Document," Report no. ASDI-FD-001, Cambridge, Massachusetts.
- Yankey, M., 2003, "User's Guide for the National Airspace System (NAS) Operational Map Display (NOMAD)," MP03W0000060, The MITRE Corporation, McLean, Virginia.

Carnegie-Mellon University. He can be contacted by email at <lwojcik@mitre.org>.

DISCLAIMER

The contents of this document reflect the views of the authors and The MITRE Corporation and do not necessarily reflect the views of the FAA or the DOT. Neither the Federal Aviation Administration nor the Department of Transportation makes any warranty or guarantee, expressed or implied, concerning the content accuracy of these views.

AUTHOR BIOGRAPHIES

BRENDAN HOGAN is a Senior Simulation Modeling Engineer at The MITRE Corporation's Center for Advanced Aviation System Development (CAASD). He has a B.S. in Mathematics and Physics from St. Lawrence University and a M.S. in Computational Operations Research from The College of William and Mary. He is a member of INFORMS. He can be contacted by email at <bhogan@mitre.org>.

LEONARD A. WOJCIK is a Project Team Manager at MITRE CAASD, where he leads a group that does system-level modeling and simulation of the National Airspace System. He has worked at MITRE for 22 years. He has a B.A. in Physics and Mathematics from Northwestern University and an M.S. in Physics from Cornell University. His Ph.D. is in Engineering and Public Policy from