MTR090220

MITRE TECHNICAL REPORT

MITRE

Advanced TFM Congestion Management Performance Analysis and Research Results

Fiscal Year 2009

Norma J. Taber
Claude K. Jackson
Anthony J. Masalonis
Lixia Song
Christine P. Taylor
Craig R. Wanke
Stephen M. Zobell

September 2009

CAASD

MTR090220

MITRE TECHNICAL REPORT

MITRE

Advanced TFM Congestion Management Performance Analysis and Research Results

Fiscal Year 2009

Norma J. Taber
Claude K. Jackson
Anthony J. Masalonis
Lixia Song
Christine P. Taylor
Craig R. Wanke
Stephen M. Zobell

September 2009

Sponsor: The Federal Aviation Administration
Dept. No.: F045
Project No.: 0209FB08-CE
Outcome No.: 8
PBWP Reference: 8-2.1-2
–Advanced TFM Congestion Management Performance Analysis and Research Results"

For Release to all FAA
Approved for public release; distribution unlimited.

Abstract

This document describes the results of Fiscal Year 2009 (FY09) research on advanced congestion prediction and automated en route congestion resolution. An improved model for aggregate traffic demand prediction uncertainty was completed. Three different models for estimating the impact of weather on sector capacity were compared, and a hybrid solution is proposed. A new technique for developing rerouting options was developed for application to near-term, semi-automated congestion management tools. An existing sequential decision-tree approach for tactical, probabilistic congestion management was converted to a continual approach that can be realistically applied to real-time decision support. Finally, two methods for improving the performance of automation-developed congestion resolution maneuvers were studied: a partial-optimization approach that can consider multiple congestion resolution goals, and an approach for adapting to poor forecasts by explicitly planning deferred resolution maneuvers.

Keywords: traffic flow management (TFM), NextGen, congestion management capabilities, weather, decision making, resolution maneuvers, benefits analysis, traffic simulation model

Acknowledgments

The authors would like to thank the support staff who helped prepare this document for publication: Cheryl Mayson-Shaffer for applying her graphics magic, Angela Signore for final formatting and standards checking, and Joan Arnold for shepherding it through the delivery process.

Table of Contents

List of Figures

List of Tables

1 Introduction

En route airspace can become congested through either excessive demand or capacity reduction, the latter often due to convective weather. Federal Aviation Administration (FAA) traffic managers in the present-day U.S. National Airspace System (NAS) control congestion primarily through manual processes, relying on experience and limited traffic prediction data to develop resolution strategies [FAA, 2006]. Traffic managers must identify and resolve impending congestion based on uncertain predictions of weather and traffic demand. They must estimate the future loss of airspace capacity due to predicted weather, and implement strategies to restrict demand such that congestion is avoided. This process may involve hundreds of flights, and is done with little or no automated decision support. As a result, current traffic management practice is to take aggressive action at strategic timescales (2+ hours before anticipated problems), using large initiatives such as Airspace Flow Programs or National Playbook reroutes. This frequently causes unnecessary delays, since weather events often do not occur as forecasted.

The Next Generation Air Traffic Management System, or NextGen for short, is being developed to handle the increasing demand for air travel [JPDO, 2007; FAA, 2008]. Congestion management methods will need to become more precise, efficient, and adaptive to keep pace. A key element of the NextGen traffic management concept is the ability to flexibly and efficiently resolve congestion at tactical timeframes (say, 30 minutes to 2 hours before anticipated congestion) by modifying individual flight trajectories. With better tactical traffic management capabilities, less intrusive measures are needed at strategic timeframes. But this kind of tactical congestion management requires automation support that can explicitly account for uncertainties in traffic and weather forecasts, and that can rapidly create flight-specific resolution actions which manage the risk of future congestion to acceptable levels.

To help define and develop such a toolset, The MITRE Corporation's Center for Advanced Aviation System Development (CAASD) undertook the work discussed here. Some of the research topics, including probabilistic demand and capacity forecasts and the reroute option development method, may be candidates for implementation in Traffic Flow Management System (TFMS) Work Package 2 (WP2) or WP3. The rest are intended to help define the eventual NextGen congestion management capability.

1.1 Tactical Congestion Management Concept

We envision real-time decision support for tactical en route congestion problems which explicitly accounts for uncertainty and adapts to changing conditions as problems evolve over time. The concept is illustrated in Figure 1-1 as a control loop, where congestion management decisions are made continually at regular intervals (e.g., every 15 minutes). The distribution of traffic demand in en route sectors is predicted based on flight plans (or down-linked aircraft data), track data, wind forecasts, aircraft performance data, and other adapted elements. Convective weather forecasts, which will eventually include measures of forecast uncertainty, are used to predict the probabilistic capacity of en route sectors. This calculation may factor in the predicted traffic demand, since the true capacity of sectors is sensitive to the traffic flow

patterns and how they interact with the weather. The distributions of demand and capacity are convolved to produce a probabilistic congestion forecast, where congestion is simply defined as when demand exceeds capacity.

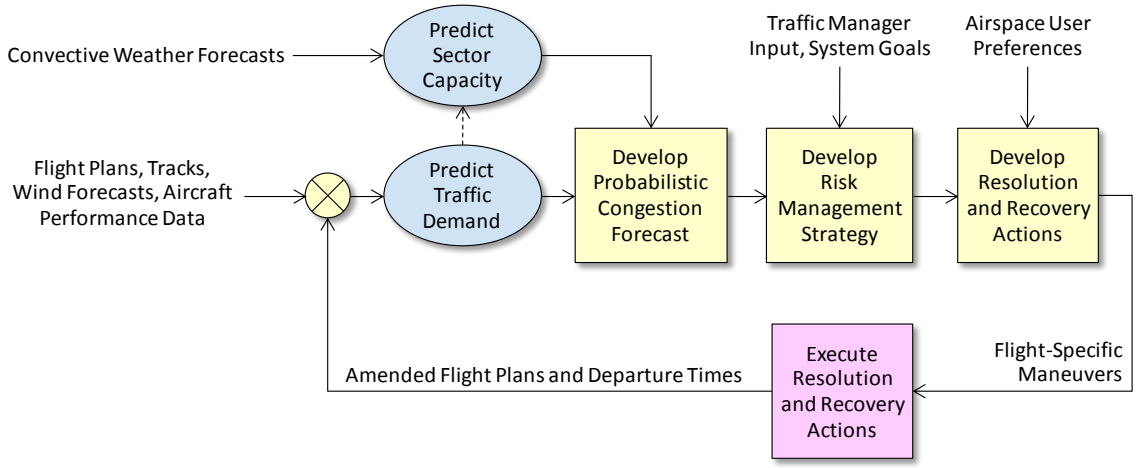

Figure 1-1. A Continual, Probabilistic Congestion Management Concept

The primary strategy decision is how aggressively to act to control predicted congestion. This is described here as the ―risk management strategy," and the best strategy depends on several things: (1) the system goal, defining the tradeoff between the risk of unsolved congestion and delay costs, (2) the congestion resolution options available, and (3) the uncertainty in the congestion forecast. Factors (1) and (2) can be adjusted by the traffic manager if desired, while (3) is a function of the traffic demand and weather forecasts used to predict congestion. The risk of unsolved congestion can be computed from the probabilistic forecasts, and the maximum tolerable risk can vary as a function of the time remaining to the predicted congestion.

Once a risk management strategy has been chosen, flight-specific maneuvers are developed to implement it. If the weather turns out to be less disruptive than predicted, delay recovery actions to undo previous maneuvers may be needed. In both cases, it is anticipated that relatively few flights would be maneuvered at any single decision time, as compared to the large-scale traffic flow initiatives commonly used today. At this step, NAS customers can collaborate with traffic management providers to coordinate resolution or recovery actions with their business needs. This may be via customer preferences, or eventually via a 4-dimensional (4-D) trajectory negotiation process. The final step is to execute the actions such that departure times and cleared flight plans, or the agreed-upon 4-D trajectory, are updated.

1.2 Research Overview

This document describes the results of Fiscal Year 2009 (FY09) research on advanced congestion prediction and automated en route congestion resolution performed under our FAA tasking. The research topics discussed in this document can be related directly to the concept elements shown in Figure 1-1. Sections 2 and 3 describe work on predicting sector demand and

capacity, the blue ovals in the figure that are used to develop the probabilistic congestion forecast. Section 4 documents an initial algorithm to generate ad-hoc reroute alternatives, which aid in developing resolution and recovery actions. Section 5 describes research on feasible risk management strategies for automated congestion resolution, and Section 6 includes new results on partial-optimization strategies for choosing congestion resolution maneuvers. Section 7 provides a summary of major findings.

1.3 Related Work

In addition to the direct FAA-funded work described here, there is a highly relevant FAA Mission Oriented Investigation and Experimentation (MOIE) project called "Integrated Departure Route Planning" (IDRP). The IDRP project aims to provide decision support for departure routing in the presence of weather and is being conducted in collaboration with Massachusetts Institute of Technology Lincoln Laboratory (MIT/LL). This effort will have direct FAA funding in FY10, aimed at developing a prototype capability for field evaluation in FY11. Details on IDRP can be found in [Masalonis et al., 2008].

Also, a MITRE-Sponsored Research (MSR) project titled "Flight Option Generation for NextGen Automation" began in FY09. It is aimed at providing a general-purpose capability for developing feasible, operationally-acceptable alternative trajectories ("flight options") for flights involved in congestion or weather-related problems. Options will be evaluated against a wide variety of efficiency and operational metrics, including: imposed operating cost, severe weather proximity, induced air traffic control complexity, sector congestion, coordination workload, and others. The work described in Section 4 is an early consequence of this research.

2 Demand Prediction

To explicitly consider uncertainty in congestion management, we need quantitative models of the uncertainty in demand and capacity predictions. This section addresses improvements made in ―aggregate demand modeling," which provides probabilistic estimates of the traffic demand on sectors, as measured by the peak flight count[1] during each 15-minute period. An initial Aggregate Demand Model (ADM) was developed in 2005 [Zobell et al., 2005; Wanke et al., 2005a], and has been used successfully for a number of studies, but has some known weaknesses. Thus we have been developing an improved version, known as ADM2, and completed this task in FY09. This section describes the process of extrapolating the ADM2 error distribution curves to higher demand cases, as well as some associated validation activities.

2.1 Introduction

In the FY08 analysis reported in [Wanke et al., 2008a], ADM2 was created using a multiple regression approach based on a training dataset of about 30 million observations. These observations represent predictions at various look-ahead times (LAT, the number of minutes between when the prediction is made and the beginning of the time interval being predicted) for all NAS sectors in the contiguous U.S. for 28 different 24-hour periods in March through June 2006. Those 28 periods were chosen because they had low prevalence of delays and convective weather. The results were validated against a test dataset, consisting of another 28 different 24-hour periods in March through June 2006, selected so as to exhibit similar characteristics (e.g., delays, weather, day of the week and month of the year) to the training data.

A separate regression equation to predict peak flight count was derived for each LAT at 15-minute intervals from 15 to 270. The equation was based on the input parameters of total number of Scheduled, Filed, and Active flights expected to traverse the sector in the 15-minute time interval of interest and in the preceding and following time intervals. (Preceding intervals are represented by appending ―P" to the variable names, and following intervals by ―N" for ―Next.") For each LAT, a weight (multiplier) was derived for each of these nine parameters (Scheduled, Filed, and Active totals for the time interval of interest, and for the intervals preceding and following it) as well as a constant term that was added to account for flights not projected to be in the sector at the given time interval that nonetheless appeared in the sector, such as pop-ups[2] and reroutes. The result of the equation represents the prediction of the peak aircraft count for the 15-minute period of interest.

---

1 Peak flight count is the greatest number of flights present in the sector in any one minute.

2 Pop-ups are unscheduled and unfiled flights not known to the system at the time the prediction is made, that appear between the time the prediction is made and the time being predicted.

As an example, the equation for LAT=15 model (with parameters rounded to two decimal places for display, though in reality greater precision is used) is as follows:

$$\begin{aligned} \text{Peak} = {} & 0.06(\text{Scheduled}) + 0.18(\text{Filed}) + 0.47(\text{Active}) \\ & + 0.05(\text{ScheduledP}) + 0.12(\text{FiledP}) + 0.09(\text{ActiveP}) \\ & + 0.02(\text{ScheduledN}) + 0.07(\text{FiledN}) + 0.01(\text{ActiveN}) \\ & + 0.46 \end{aligned} \qquad \text{Equation 2-1}$$

Next, error distribution curves were generated for each combination of LAT, Adaptation Controlled Environment System (ACES) altitude band (Low, High, or Super-high), and ~~"N~~," which represents the peak count point estimate produced by the regression equation (referred to as ~~"P~~eak" in Equation 2-1).[3] Error is defined here as the integer number of flights by which the peak count was over predicted or under predicted. The error curves were actually probability mass functions (PMFs) for the likelihood of each integer error value given the Altitude-LAT-N combination. In this section, the PMFs are represented as probability distribution functions (PDFs) for ease of visualization and of comparing different distributions. In [Wanke et al., 2008a], the error distributions observed in the training data were used as the probabilities in the final distribution model. For the low to moderate values of N (up to 12) used in [Wanke et al., 2008a], this approach was found to be very accurate, as the probabilities matched those in the test dataset very well.

For higher values of N, we determined that the training set probabilities were not appropriate for use in the model. Comparing Figure 2-1 and Figure 2-2 helps illustrate this. Each of these two figures represents how the error distribution varied with N, averaged across all sector types and LATs (again, for the model, separate distributions were generated for each combination of Altitude, LAT, and N). Positive numbers on the horizontal axis represent over prediction.

[3] This count represents the maximum number of flights in a single sector within any one minute during the 15-minute time bin of interest.

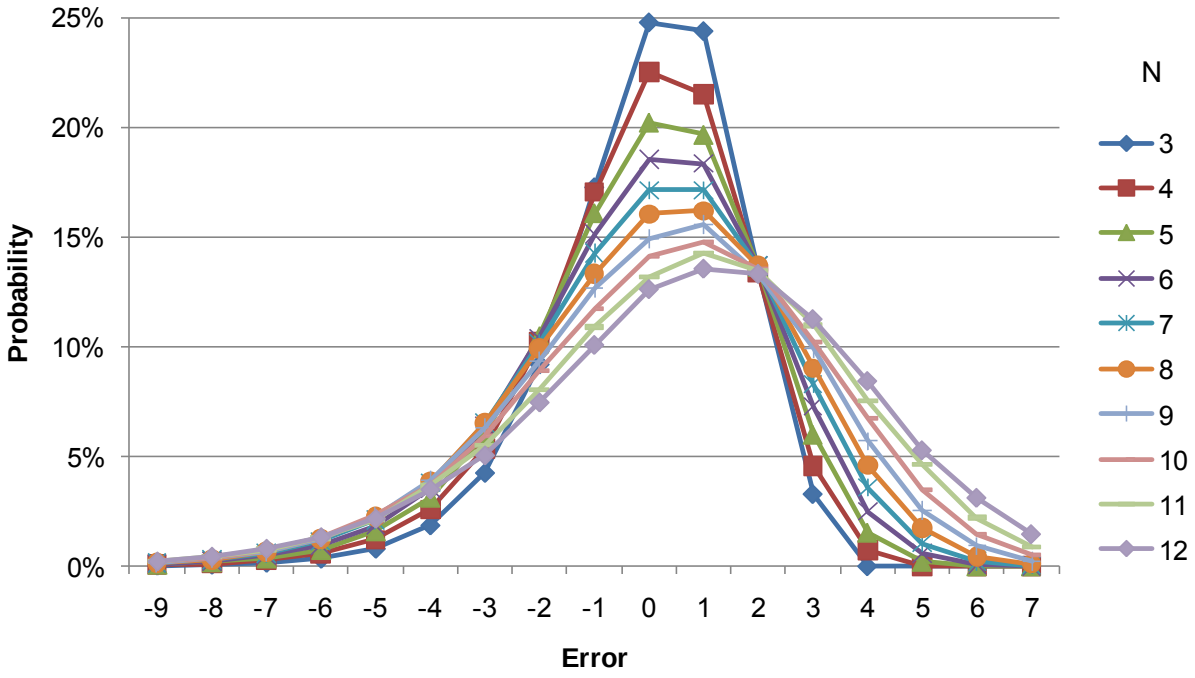

Figure 2-1. Error PMF for Each N in [3, 12], Averaged Across All Altitude and LAT

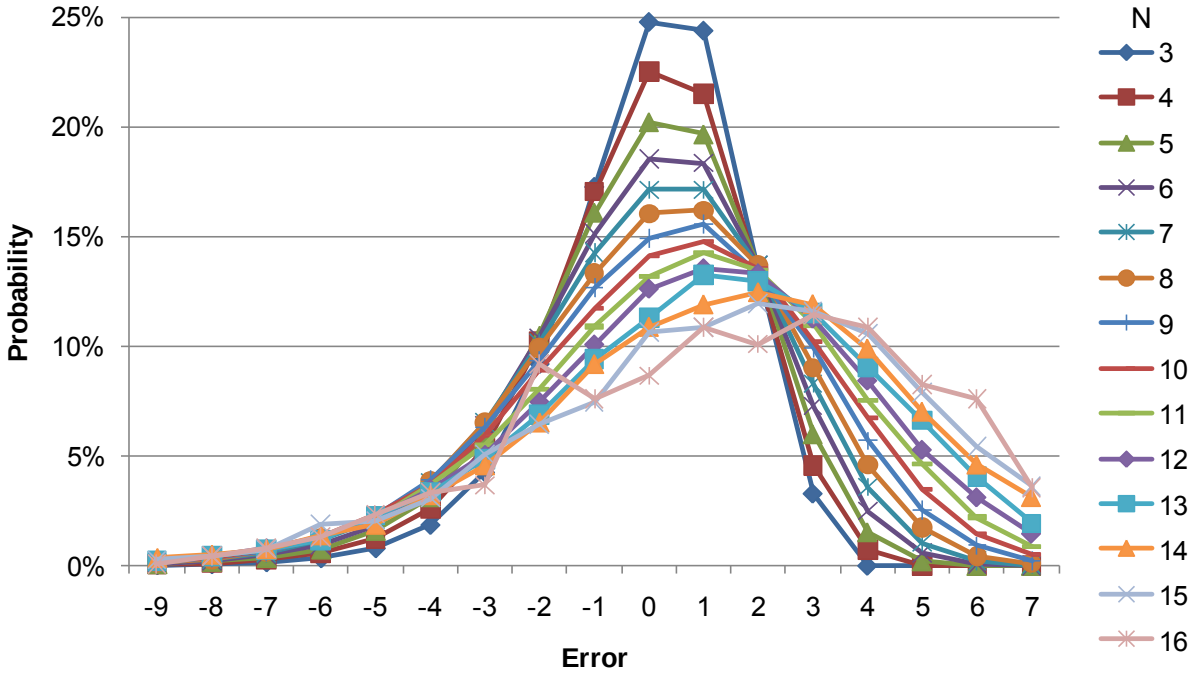

Figure 2-2. Error PMF for Each N in [3, 16], Averaged Across All Altitude and LAT

In Figure 2-1, the progression moves to the right and becomes flatter (higher dispersion) as N increases because when there are more flights believed to be coming, there is more uncertainty that the predicted number of flights will be the exact number. Over prediction is more likely due to the chance that some of those flights expected will not appear in the sector at all, or at least not during the time when the peak count situation is present.

As can be seen in Figure 2-2, when larger N's are used, the curves become erratic; this occurs for two related reasons. First, in higher demand cases there is the increased possibility that actions by traffic flow management (TFM) or other air traffic management (ATM) personnel have made the actual peak count lower than the true demand on the sector. An attempt was made to reduce the effects of this confounding factor by filtering out cases where the predicted demand, relative to the Monitor/Alert Parameter (MAP) threshold, was high enough that ATM actions might be likely.[4] However, the possibility remains that relatively high demand—even if not very close to MAP—may have motivated ATM personnel to take action to reduce the demand.

The second reason for the erratic shapes of the higher-N curves is that the results for high-N cases do not come from a large enough and representative enough sample of the data. A smaller number of different sectors are present in the data, for reasons directly related to the filter described in the previous paragraph. This approach enhanced the results' validity overall, but eliminated a large percentage of the data for higher-N cases in sectors having lower MAP values. This phenomenon is discussed in more detail in Section 2.4.2. In general, the high-demand filter reduces the total number of observations for high-N cases, which are already of limited prevalence since higher-demand situations are rarer overall than lower-demand situations.

Several different methods were attempted to derive smoother and more reliable estimates for the PMF functions at these higher N's. This section will report the two primary methods analyzed: linear extrapolation of the PMF probability values as a function of N, and the method that was ultimately used, a two-step process involving curve fitting to the PMFs and extrapolation of the curve parameters.

---

[4] The MAP value (see further discussion in Section 5.1) is the peak sector count at which investigative or ameliorative action is expected to be taken. Analysis conducted for the initial ADM [Wanke et al., 2005a; Zobell et al., 2005] shows evidence that TFM actions to reduce the demand may begin to occur when the predicted peak count is above, at, or within 6 flights below MAP. Therefore, cases where the predicted demand was greater than (MAP-6) were filtered out of analysis for the present study. For this filter, predicted demand was defined as the peak count prediction of the deterministic model against which the ADM2 is compared—that is, the predictions made by the CAASD Collaborative Routing Coordination Tools (CRCT) prototype, which correspond closely to the Enhanced Traffic Management System (ETMS) predictions that operational personnel would have seen and based decisions on during the days when the data were collected.

2.2 Extrapolation of Raw PMF Percentage Along N

2.2.1 Method

Visual examination of the data from the FY08 analysis suggested that the probability of a given error value (for example, Error=-1) in the PMF for a given combination of Altitude, LAT, and N, was approximately a linear function of N, at least in a certain range of N, within that Altitude-LAT combination. For example, Figure 2-3 shows data only for the Altitude-LAT combination of High sectors and LAT=270, and plots all the probability values as a function of N for each error value from -5 to 5.

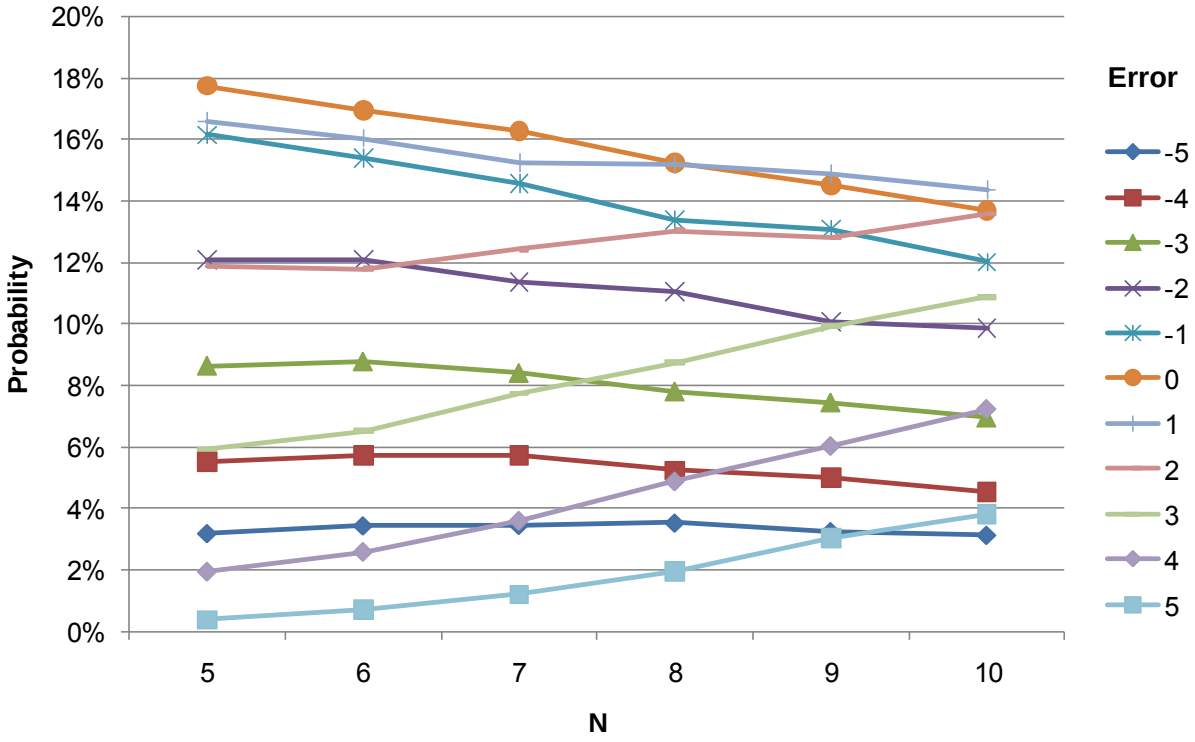

Figure 2-3. Probability of Error as a Function of N for Errors in [-5, 5], High Sectors, LAT=270

This plot contains the same data as would a set of PMFs, one for each N, for this specific Altitude-LAT combination, except that here, the variables defining the horizontal axis and the separate data series (Error and N, respectively, in a multiple PMF plot such as Figure 2-1 or Figure 2-2), are switched. To better understand what the figure depicts, consider the points where the horizontal axis (which represents N) is 5 (the leftmost part of the graph). The series for Error=0 shows that where n=5, there is approximately an 18% chance the prediction of 5 is exactly right. The curves below it show about a 16% chance each of an error of 1 (over prediction of 1: peak equals 4) and an error of -1 (under prediction of 1: peak equals 6). The sum of the values of each of the Error series in [-5, 5] at any given N, that is, any given value on the

horizontal axis, will be approximately 1, because the vertical axis values on the graph represent the probability of that error value given the combination of Altitude=High, LAT=270, and N=5. Technically, because errors of less than -5 and more than 5 are possible, the sum of the values shown here will be slightly less than 1. Errors outside this range (all errors from -9 to 7 were used in the model for reasons explained in [Wanke et al., 2008a]) are always very small and would only add clutter to the graph.

Figure 2-3 shows a key general trend in the data: error values close to zero, especially under predictions, decrease as N increases. Larger absolute error values, especially over predictions, tend to increase with N. This is why some of the curves (lines) slope upward and some downward.

Through visual examination and analysis of how well linear functions could be fit to the data series depicted in Figure 2-3 and the analogous series for all combinations of Altitude, LAT, and Error, it was determined that the best range of N from which to linearly extrapolate would be 5 to 10. Models were run to fit a line to these curves. After deriving the best-fitting linear function of N for each combination of Altitude band, LAT, and degree of error, the lines were extrapolated beyond N=10 to higher numbers. After the probabilities were derived for each combination of Altitude, LAT, N, and Error, the small probabilities outside the range of error values in [-9, 7] were discarded. The PMF for each Altitude-LAT-N combination was normalized so that the sum of the error probabilities for this range of errors equaled 1.

2.2.2 Validation

Validation was conducted by first determining the goodness of the linear fits to the probabilities as a function of N as described in Section 2.2.1, and then assessing how accurate and logical the resulting extrapolated probabilities were.

First, the goodness of fit of the linear functions was assessed using linear regression, which determines how well a straight line can be fitted to a series of data points; the measure of linearity we employed is known as R^2. R^2 is used for various analyses throughout this document and is one of several possible measures of how well the points in a series of data fit some hypothesized function, such as a straight line or a normal curve. The R^2 statistic, more generally speaking, represents the proportion of the variance in the variable(s) of interest that is explained by the model as opposed to resulting from random error. Being a proportion, its maximum possible value is 1.0. There is no single criterion for a "good" R^2; this depends among other things on the size of the dataset and the relative importance of statistical significance and effect size in the application being studied. In the analyses reported in this section, values above 0.9 can generally be considered to reflect a high degree of confidence that the data really conform to

the line or curve being assessed. Note that all R^2 values reported in this section are actually adjusted R^2 values.[5] However, they are generally referred to in the text simply as R^2.

In the first step of the extrapolation process, an R^2 was derived for each series of probabilities for N=5 through 10. This was done for each of 918 combinations of Altitude (3: Low, High, Super-high), LAT (18: {15, 30, … 270}), and Error (17: {-9, -8, … 7}). The data for Low sectors, LAT=15, and Error=-9, did not have data points for any N in [5, 10] (i.e., these combinations were never observed in the training data), and for Super-high sectors, LAT=15, and Error=-9, there were only two data points with N in [5, 10], making it impossible to estimate the goodness of fit of a linear function to the series. Thus, a total of 916 data series were fitted to linear functions. Most (800) of these series contained 6 data points, one for each N in the range analyzed. The remainder had at least one N with no data and hence contained between 3 and 5 points. The distribution of the R^2 is shown in Figure 2-4, with the bars representing the percentage of the 916 calculated R^2 that were less than or equal to the value indicated on the horizontal axis and greater than the next value down, and the line showing the cumulative percentage of cases less than or equal to the horizontal axis value. Almost 29% (263) of the linear regression analyses have an R^2 of less than 0.7.[6]

---

[5] The adjusted R^2 statistic controls for the number of variables in the model and is recommended when the analysis is based on a sample, rather than on the entire universe of data to which the model is being applied. Adjusted R^2 is always less than or equal to R^2 and therefore is generally a more conservative estimate of fit.

[6] The selection of 0.7 as upper boundary of the first bin of the histogram in Figure 2-4 is not an indication that an adjusted R^2 below 0.7 necessarily represents a poor fit. In fact, most of the series in the present data that exhibited R^2 in this vicinity have statistically-significant linear fits, meaning that the data points are unlikely to be arranged in approximately a straight line due to random chance. However, the fits of these series should be viewed relative to the far better fits observed using the modeling methods described in later sections.

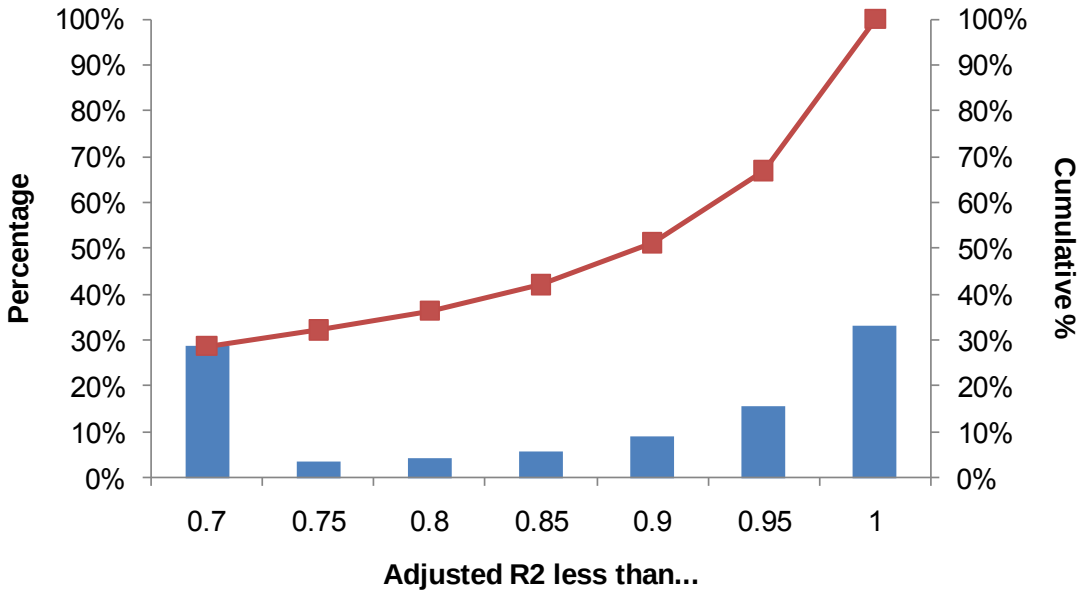

Figure 2-4. Distribution of R^2 for Error Probability as a Linear Function of N for All Valid Combinations of Altitude, LAT, and Error

Next, the lines were extrapolated beyond N=10 to higher numbers while forcing the sum of the probabilities across error values for any Altitude-LAT-N combination to be 1. Unfortunately, when this was done, many illogical values were derived. The reasons for this can be understood by referring back to Figure 2-3 and the associated text. Prediction errors near zero decrease as N increases, and larger errors, especially over predictions, increase with N. If the linear extrapolation is extended to a large enough value of N, inordinately high and low values are inevitable.

Figure 2-5 illustrates example extrapolations for one combination of Altitude and LAT. The values for N in [5, 10] are the actual probabilities observed in the training dataset and are identical to what is plotted in Figure 2-3. The points making up the rest of each line are the result of the linear extrapolation.

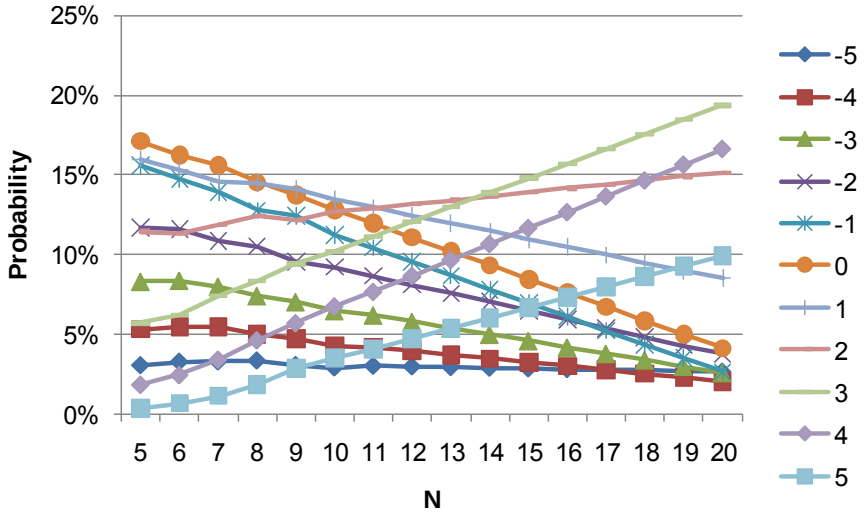

Figure 2-5. Linear Extrapolation of Error Probability as a Linear Function of N, High Sectors, LAT=270

In the extrapolated data, the probability of the most accurate predictions continues of course to go down with N and the probability of the less accurate, especially over predictions, goes up. Note that as in Figure 2-3, to avoid clutter, this and some of the other figures in this subsection show only errors between -5 and 5. However, in the modeling, errors from -9 to 7 were still used.

When these data are converted back into PMFs with Error on the horizontal axis, the curves formed by the extrapolations to some of the higher N values take on odd shapes and depict unusually high percentages of over prediction, owing in part to Error=3 (see Figure 2-5) having high values and a rather steep upward slope as N approaches 10. Figure 2-6 shows the same data as Figure 2-5, converted to PMFs (again, the data for N values between 5 and 10 are the original probabilities, while extrapolated probabilities are used for N values of 11 through 20).

Extrapolating further out, the values for parts of the curves become not only unusual but impossible, with negative values appearing, as can be seen in Figure 2-7.

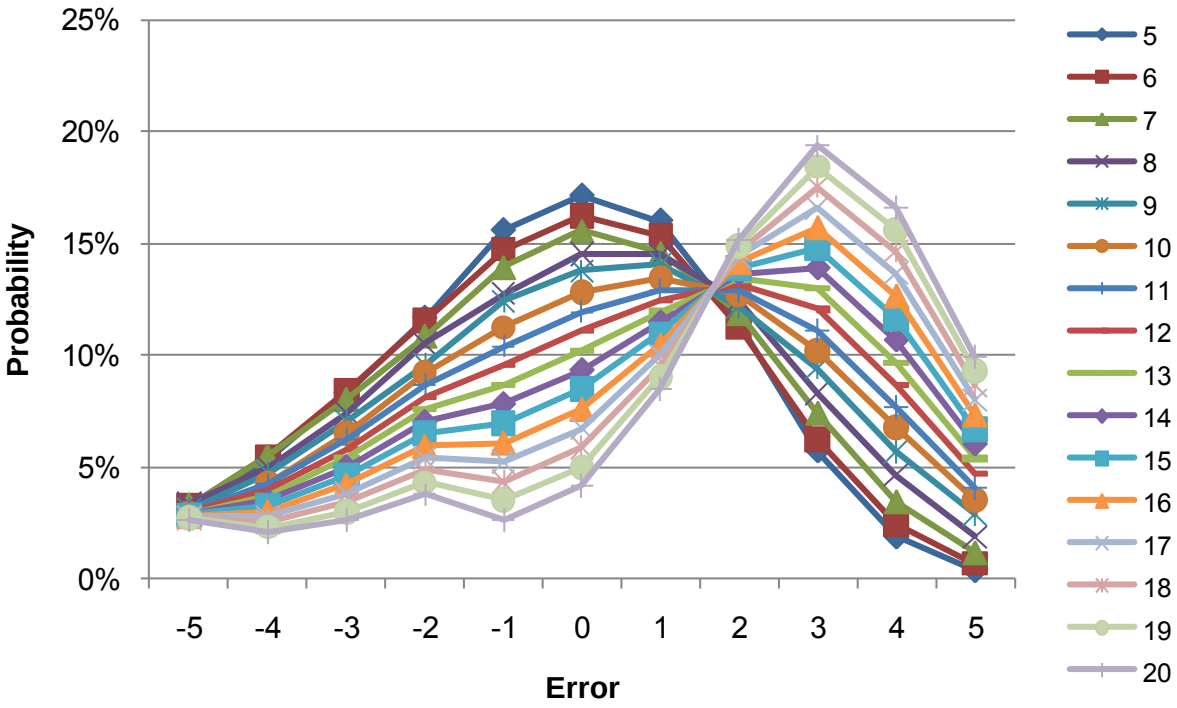

Figure 2-6. Error PMF for Each N in [5, 10] (Training Data) and [11, 20] (Linear Extrapolation of Probability along N), High Sectors, LAT=270

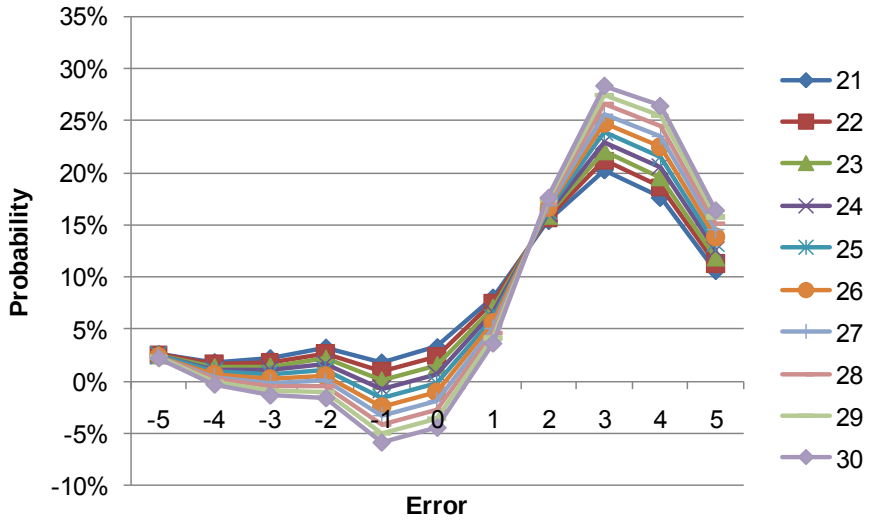

Figure 2-7. Error PMF for Each N in [21, 30] (Linear Extrapolation of Probability along N), High Sectors, LAT=270

While the above figures and descriptions present the results for High sectors at LAT=270 as an example, similar oddities were observed for most combinations of Altitude and LAT.

Of course, additional correction factors and constraints could be applied to ensure that the PMF probability values were always greater than 0 and to smooth out the illogical inflections in the curves. However, such methods would require additional assumptions and manipulations that might render the extrapolation less reliable. Therefore, despite the initial appeal of capitalizing on the linear nature of the relationship between N and error probability for many Altitude-LAT-Error combinations, it was decided to pursue other methods for extrapolating to high-demand (high N) cases.

2.3 Curve Fitting and Extrapolation of the Distribution Parameters Along N

2.3.1 Method

Another approach was attempted, based on the approach used for the initial ADM. In that model [Zobell et al., 2005; Wanke et al., 2005a], the distribution was created by fitting common distributions to the error PMFs for each combination of the conditioning variables of interest. Various distributions were attempted and the final model used Poisson distributions for some datasets and binomial for the others. A similar curve fitting approach was attempted for the PMFs in the present model.

For this analysis, a two-step process was followed. The first step was to fit a function to the training data probability values and save the parameters describing that function.[7] Next, separately for each combination of Altitude and LAT, the distribution parameter values as a function of N were fit to other appropriate distribution types (mostly linear functions) and these latter distributions extrapolated to higher values of N. This method will be described in more detail in Section 2.3.2 by showing for one subset of data the derivation and validation of the distribution type ultimately chosen for the model, the Gaussian (normal) distribution.

A number of different distribution types were modeled, including Weibull, Gaussian, Johnson Su and Sb, and others. The data series for the curve fitting was the PMF representing the probability of each error value for a given combination of Altitude, LAT, and N. A total of 540 curves were fit (3 Altitudes x 18 LAT x 10 N), representing the same data ~~"b~~ins" used in [Wanke et al., 2008a]: Low, High, and Super-high Altitudes, LAT from 15 to 270 in 15-minute intervals, and N from 3 to 12. It was found that the Gaussian curves fit the PMF data very well across a wider range of bins than other types of distributions, and offered the additional advantage of having easily interpretable parameters of mean and standard deviation (SD) that increased with N in a nearly linear fashion.

---

[7] Different distribution types are defined based on different sets of parameters. Distribution parameters generally include values measuring central tendency, dispersion, curve shape, and other features.

2.3.2 Validation

This section describes the validation of the Gaussian curve fitting parameter extrapolation, along with further detail about how it was derived and an example thereof. As described in Section 2.3.1, other distribution types were investigated as well, and did not perform as well as Gaussian distributions according to the validation techniques discussed in this section.

First we determined how well the PMFs of the measured error distributions (see for example Figure 2-1 and Figure 2-2) fit a Gaussian function. This is not an indication of prediction accuracy; it is an assessment of how smoothly the errors are distributed rather than how large or small the errors are. For each of the 540 combinations of 3 Altitude, 18 LAT, and 10 N, we derived the best-fitting Gaussian curve to the 13 to 17 available data points.[8] Regression analysis was used to assess how well each Gaussian curve fit the 13 to 17 points in the PMF data series. Figure 2-8 depicts the distribution of the R^2 values in the same manner as Figure 2-4.

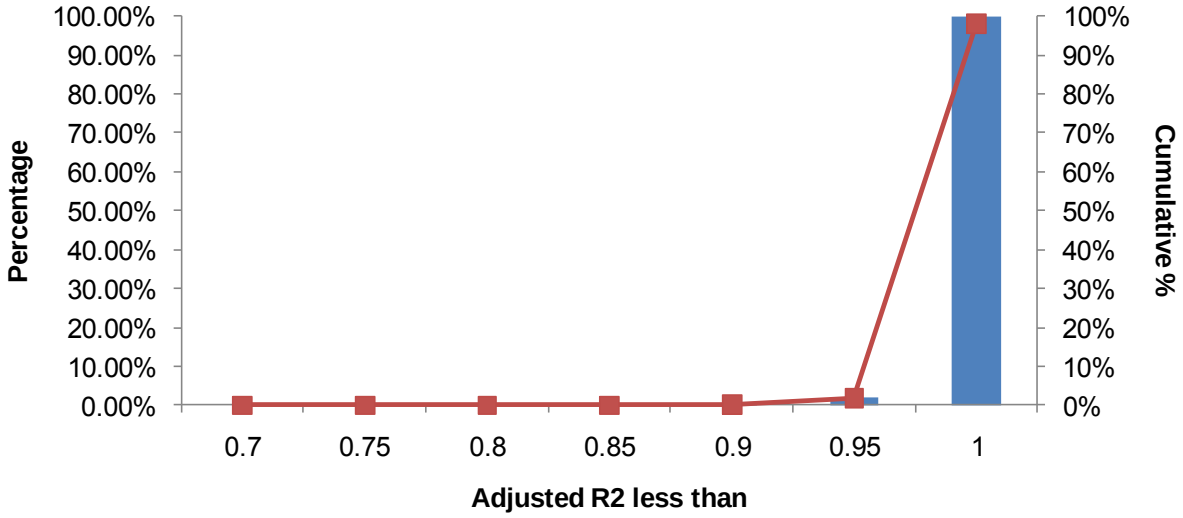

Figure 2-8. Distribution of R^2 for Gaussian Fit to Error PMFs for All Valid Combinations of Altitude, LAT, and Error

In general, it can be said that the PMF error curves fit a Gaussian distribution very, very well. Among the 540 Altitude-LAT-N combinations, the mean R^2 is 0.9846, and the minimum is 0.8997. It should be noted that all of the 15 lowest R^2 values occurred for bins where N=12. If N=12 is excluded from analysis, the lowest R^2 is 0.9610. Many of the other ~~"l~~lower" fits (in the vicinity of 0.96 or 0.97) were toward the low and high ends of the range of N's analyzed. On the other hand, 16 of the 20 highest R^2 values were observed for N in [5, 10].

---

[8] The 17 data points, where all are present, represent integer error values in [-9, 7]. Values between 4 and 7 are not available for all combinations, since N values as low as 3 were included, and the prediction error value could only be greater than N in the impossible case of negative actual peak count.

The next validation step was to examine how well the parameters of the Gaussian curves could be extrapolated beyond the data used to derive them. One issue requiring investigation was what range of N should be used as the baseline from which the extrapolation to higher N's would be conducted. As mentioned in Section 2.3.1, N values of 3 to 12 were used. The analysis in [Wanke et al., 2008a] had generated reliable results throughout this range of N, and as noted above, the fits to a Gaussian distribution were very good throughout this range. Therefore, the plan was to use the Gaussian parameters from this entire range as the basis for extrapolation. However, trial and error examination of the how the parameters varied with N suggested that [5, 10] would be preferable. The results presented in the preceding paragraph provide some support for this decision. The effects at different N values will now be illustrated further using an example subset of data bins, for LAT=270, High sectors. This combination, like every Altitude-LAT combination, initially had 10 different error PMFs, one for each N in [3, 12]. The PMFs are shown in Figure 2-9. The mean and SD of each of these fitted curves are shown in Figure 2-10 and Figure 2-11.

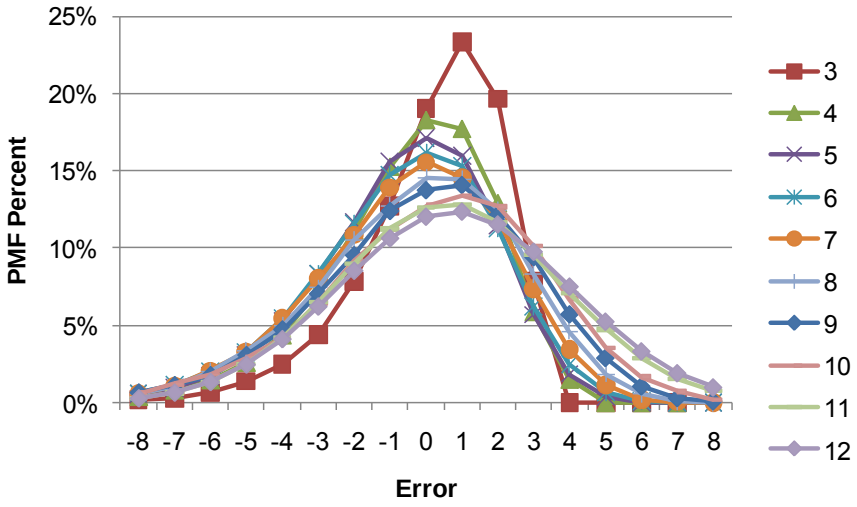

Figure 2-9. Error PMF for Each N in [3, 12], High Sectors, LAT=270

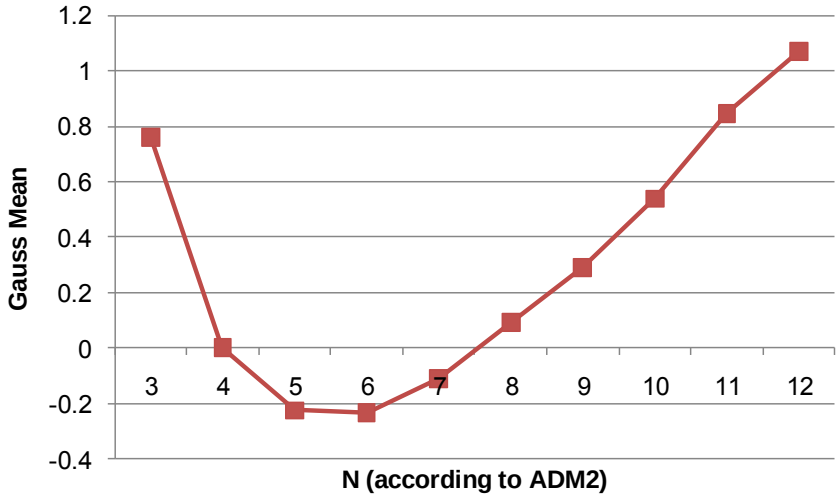

Figure 2-10. Fitted Gaussian Mean of Error PMF for Each N in [3, 12], High Sectors, LAT=270

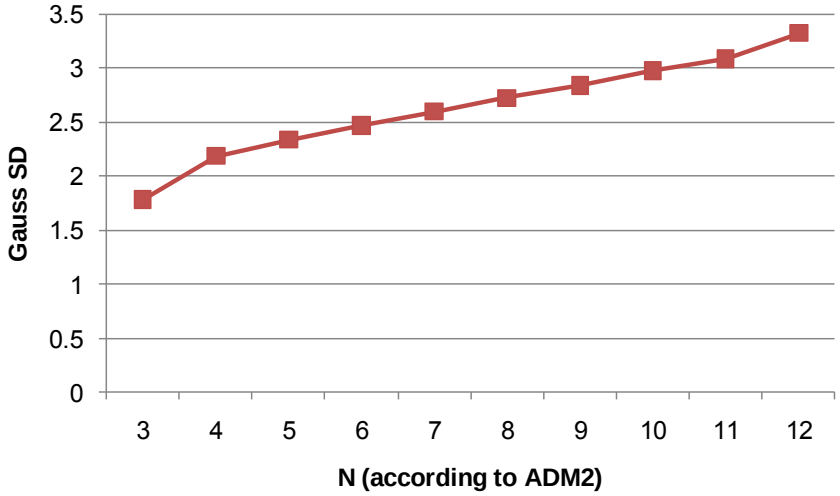

Figure 2-11. Fitted Gaussian SD of Error PMF for Each N in [3, 12], High Sectors, LAT=270

Attempts could be made to fit various types of curves to these functions; unfortunately, when looking at all combinations of Altitude and LAT, the curves showing the relationship of Gaussian mean to N and that of Gaussian SD to N were not all the same shape. However, when only N in [5, 10] were used, the functions appeared very close to linear for virtually every combination of Altitude and LAT. The data plotted in Figure 2-10 and Figure 2-11 is shown again in Figure 2-12 and Figure 2-13 for only the N in [5,10].

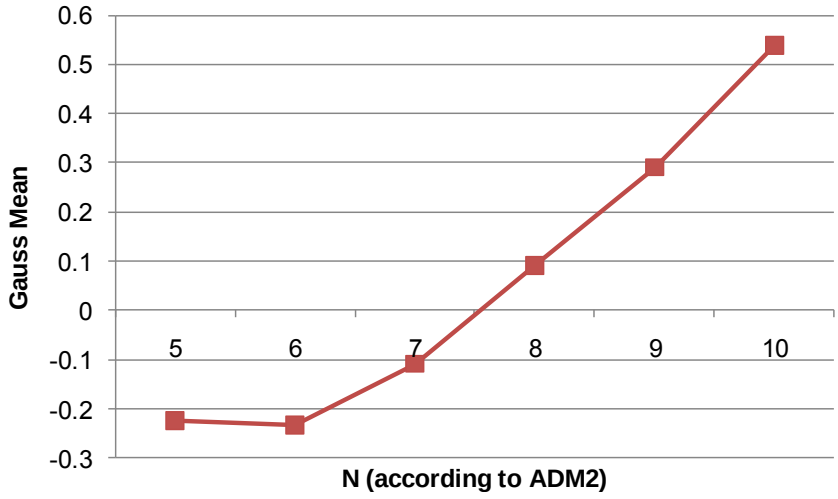

Figure 2-12. Fitted Gaussian Mean of Error PMF for Each N in [5, 10], High Sectors, LAT=270

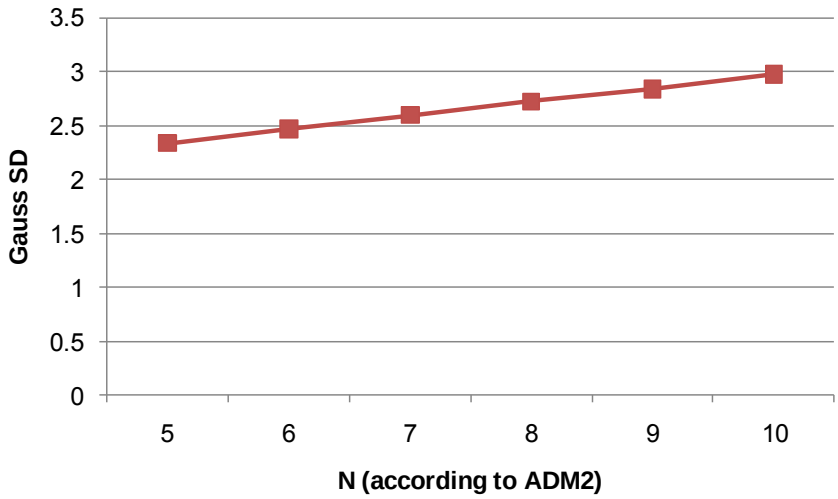

Figure 2-13. Fitted Gaussian SD of Error PMF for Each N in [5, 10], High Sectors, LAT=270

Note that it is coincidental that, for N, [5, 10] proved to be the best range to extrapolate from for both the linear extrapolation of the raw percentages described in Section 2.2 and the linear extrapolation of the Gaussian parameters detailed in this section.

Aside from the slight departure from linearity for the mean (Figure 2-12) for N values of 5 and 6, both functions appear quite linear. The R^2 for a linear fit of the 6 data points in Figure 2-12 is 0.9147. This may be considered a very good fit, and most combinations of Altitude and LAT exhibited even higher R^2 for the linear fit. For all combinations, the mean and SD both tend to increase linearly with N. This finding is consistent with the appearance of the PMFs in, for

example, Figure 2-1 and Figure 2-9, at least for N values in [5, 10] and nearby values. As pointed out earlier, it can be seen from these sets of PMFs that as N increases, the curve becomes slightly flatter and shifts slightly to the right. Also note that 270 minutes is the longest LAT used in the model, so its data can be expected to be more erratic than that for lower LAT. It is encouraging that the examples illustrated in this section used LAT=270 and nonetheless exhibit good fits both for the Gaussian distribution itself and for the linear extrapolation of the mean and SD parameters.

Linear regression analysis for all 54 Altitude-LAT combinations shows that the apparent linear nature of increase with N in the Gaussian mean and SD is borne out statistically. Figure 2-14 (for the mean) and Figure 2-15 (for the SD) show the distribution of the R^2 values of the regression equations that attempt to fit a line to the 6 points from N values in [5, 10] for each of the 54 combinations of 3 Altitude and 18 LAT. As with the R^2 analysis for the Gaussian fits, it is appropriate to reiterate here that these R^2 do *not* represent traffic prediction performance of the model. Rather, they indicate how well a straight line can be fit to approximate the data series describing the mean and SD of the fitted Gaussian curve, as a function of N, for each Altitude-LAT combination (for example, the series shown in Figure 2-12 and Figure 2-13).

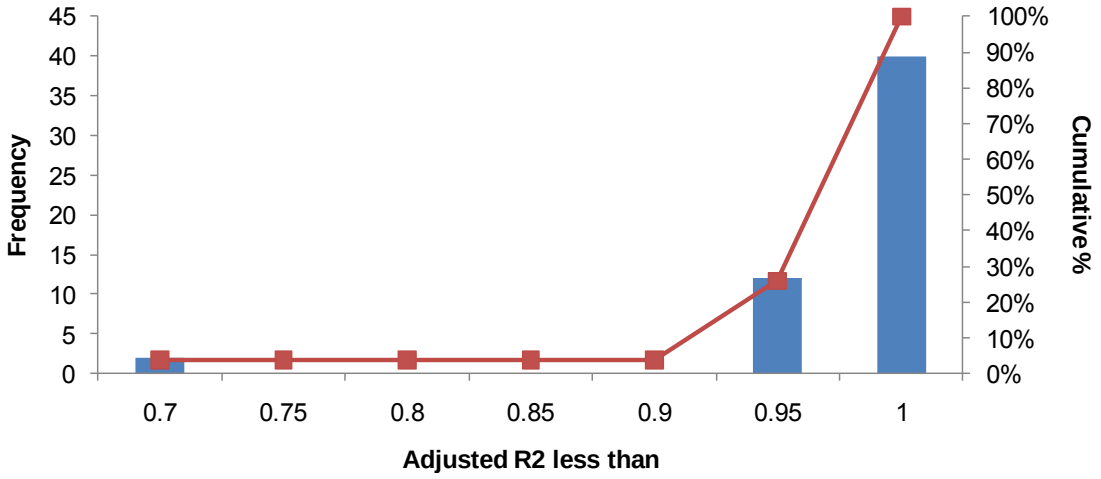

Figure 2-14. Distribution of R^2 for Linear Fit to Gaussian Mean as a Function of N, for N in [5, 10] and All Valid Combinations of Altitude, LAT, and Error

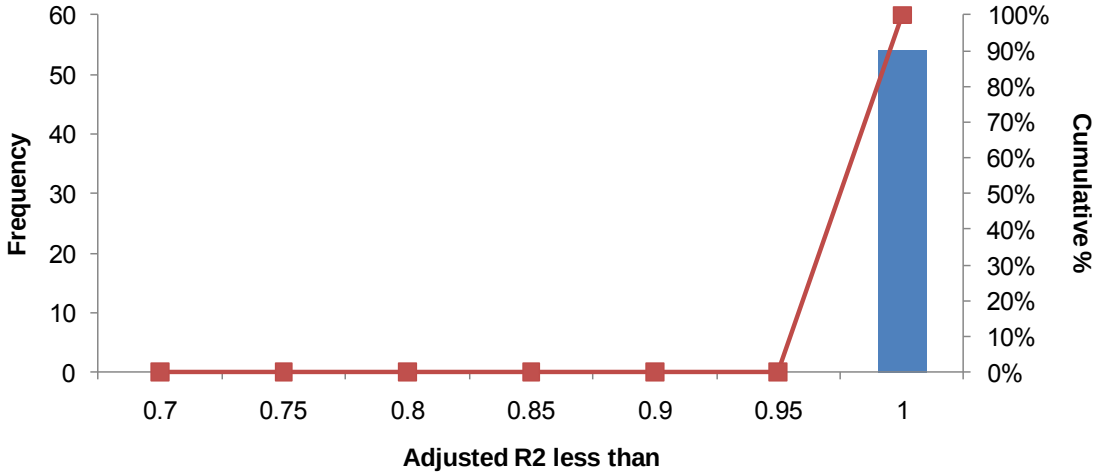

Figure 2-15. Distribution of R^2 for Linear Fit to Gaussian SD as a Function of N, for N in [5, 10] and All Valid Combinations of Altitude, LAT, and Error

The R^2 for the mean parameter is generally very high, indicating an excellent fit to a linear function. The mean R^2 of all 54 Altitude-LAT combinations is 0.9405. The two that are less than 0.7 are for Low sectors and LAT of 15 and 30. If they are excluded, the mean adjusted R^2 of the remaining 52 is 0.9655. Further investigation showed that for these two combinations, the mean of the Gaussian curve did increase with N, just like the other 52 combinations, through N=9, but then began to decrease. It is believed that this was due to more short-notice "pop-ups" occurring in Low sectors when they are already fairly busy (but not close to MAP).[9] ATM actions appear to be affecting the Low sector data at LAT=15 and 30, and N=10 (and higher). Thus the model was adjusted to use only the mean values for N in [5, 9], and for each of the two series, the extrapolation was done from these 5 data points to values of N=10 and higher. The R^2 of a linear fit to only the 5 points for each series were 0.80 and 0.93 for LAT=15 and 30, respectively.

For the SD, all 54 R^2 are above 0.95. The mean adjusted R^2 is 0.9918 and the minimum is 0.9607.

Extrapolating the mean and SD linearly from the results for N in [5, 10] to higher N's, and creating the PMFs, resulted in more logical curves, without the non-intuitive inflections and values less than zero seen in Figure 2-6 and Figure 2-7. The improvements to the PMFs effected by this method of extrapolation can be seen in Figure 2-16 and Figure 2-17, for N in [5, 20] and

[9] "Pop-ups" here does not necessarily mean pop-ups in the traditional sense of unscheduled and previously unfiled flights that appear in the system. Rather, in this situation it can and likely does include already-filed flights whose route—most likely the arrival or departure route—is altered at fairly short notice. This may be done for fix balancing purposes at times when the airports served by the sectors in question are busy. It is believed that the sectors are under predicted at higher N's because while busy, they are still at least a few flights below MAP and can accept short-notice reroutes.

[21, 30] respectively. As with most of the earlier figures and examples, these data are for High sectors, LAT=270.

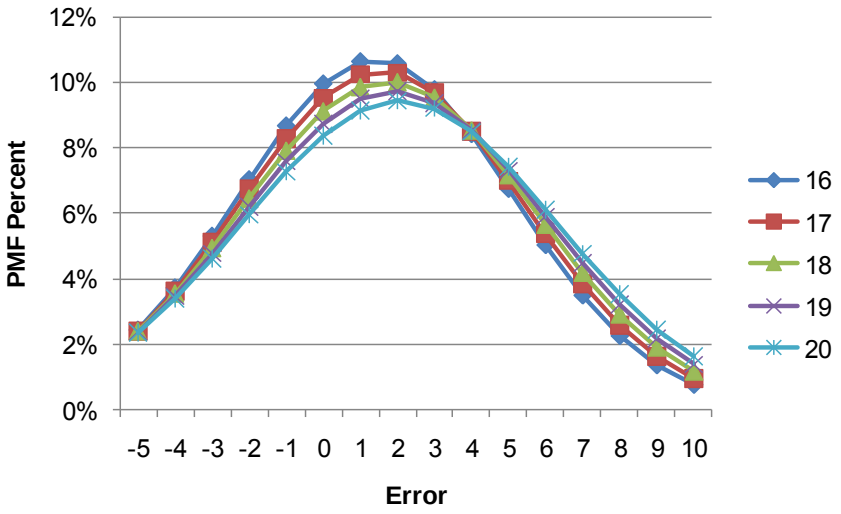

Figure 2-16. Error PMF for Each N in [5, 10] (Training Data) and [11, 20] (Linear Extrapolation of Gaussian Mean and SD along N), High Sectors, LAT=270

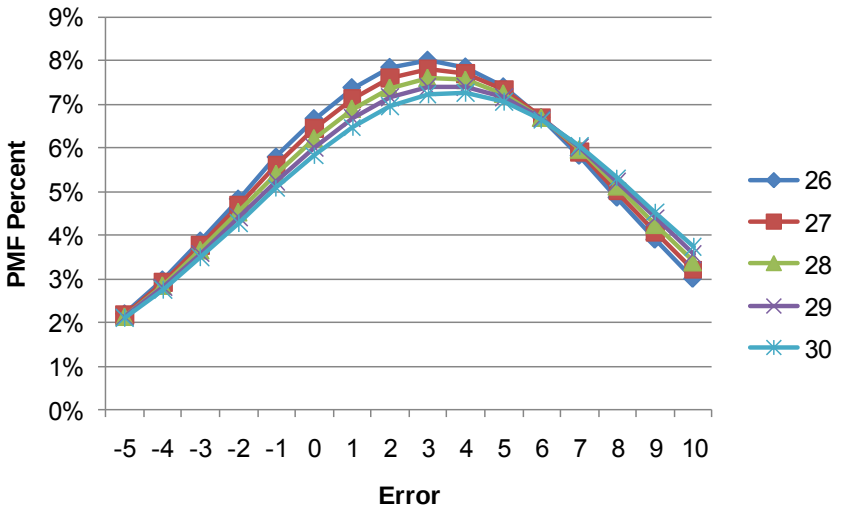

Figure 2-17. Error PMF for Each N in [21, 30] (Linear Extrapolation of Gaussian Mean and SD along N), High Sectors, LAT=270

2.4 Generating the Final Extrapolated Model

2.4.1 Method

To create the values for the extrapolated (high N) part of the model, it was necessary to determine which error values should be included. In the training data, for N values of 12 and below, error values less than -9 and greater than 7 were seen less than 0.1% of the time, and so were not used in the model for that range of N [Wanke et al., 2008a]. However, since the SD increases with N in the extrapolation, the PMFs for higher N's are more likely to contain more extreme error values. When the PMF values were determined for the higher N's, it was considered desirable to expand the range beyond [-9, 7] to include error values that were reasonably likely to occur, without including values that were so extreme as to unduly complicate the model, and that would probably occur only under abnormal operational circumstances. Therefore, a probability cutoff of 1.0% was adopted for inclusion of a given error value in the model.

For the final model as in the earlier un-extrapolated version [Wanke et al., 2008a], extreme error values were discarded and the resulting PMFs normalized in order that the error values would sum to 1 and thus represent a true PMF.

Error values from -10 to 18 were used in the new version of the model, because these are the smallest and largest error values for which probabilities of about 1.0% or higher were observed in the extrapolation for at least one combination of Altitude, LAT, and N. Errors outside the range of [-10, 18] are assumed for the model to have a probability of zero. For simplicity, this error range was used for all Altitude-LAT-N combinations, even though the modeled probabilities for the extreme ends of the range are well below 1.0% for some combinations.

2.4.2 Validation

Validation of the extrapolation focused on comparing the modeled and actual probabilities in the test dataset for each error value at each combination of Altitude, LAT, and N.

First, as an additional validity check for the Gaussian modeling of the PMFs, the probabilities calculated from the extrapolated Gaussian means and SDs were compared with the actual probabilities in the test set for the range of N—5 to 10—from which they were built. The correlation was calculated between the probability of each error value according to the Gaussian model versus the corresponding probability[10] observed in the test dataset. For this analysis, we excluded Altitude-LAT-N-Error combinations where the total number of observations[11] in the test set was less than 30, since probabilities computed based on very low numbers of observations may be skewed. There were 4774 combinations with at least 30 test set observations each. The R^2 of this correlation analysis, which assesses the extent to which the test

[10] That is, the probability of the same Error value given the same Altitude-LAT-N combination.
[11] The total number of observations is the numerator of the probability.

set probability can be predicted from the Gaussian model by a linear function, was 0.9880. Figure 2-18 shows the scatter plot of these data.

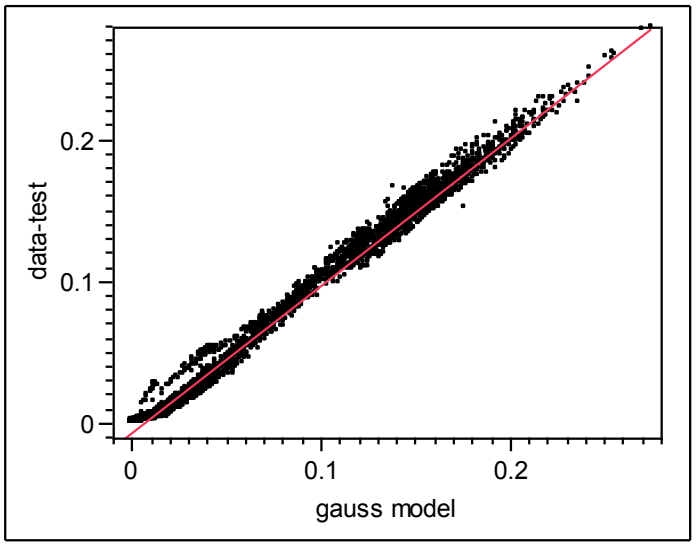

Figure 2-18. Scatter Plot of Gaussian Model and Test Data Error Probabilities for Each Combination of Altitude, LAT, N, and Error, for N in [5, 10]

This finding validates not only that the Gaussian model fit the data very well for N in [5, 10], but also that the probabilities calculated from it provide excellent prediction of the probabilities in the test set. However, for this range of N, the original method adopted for the model, where the actual probabilities observed in the training set were used directly, performed slightly better. The same 4774 Altitude-LAT-N-Error combinations as in the previous analysis were used here. Figure 2-19 shows the scatter plot; the R^2 of the correlation analysis was 0.9978.

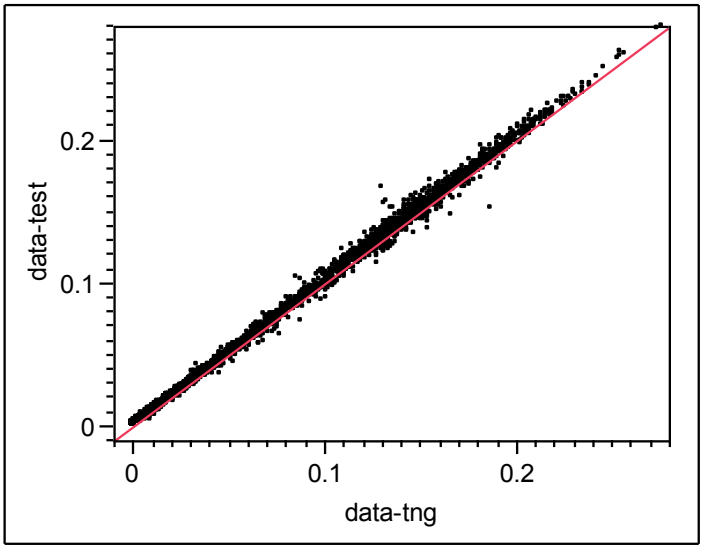

Figure 2-19. Scatter Plot of Training Data and Test Data Error Probabilities for Each Combination of Altitude, LAT, N, and Error, for N in [5, 10]

Therefore, for the final model it was decided to continue to use the training data for N in [5, 10]. The training data were also used for N in [1, 4], since the analysis conducted for [Wanke et al., 2008a], as well as additional analysis conducted this year, showed very high correlations between the training and test set probabilities for all low values of N.

To validate the extrapolated portion of the model, the probabilities derived from the Gaussian model were compared to those in the test set for N in [11, 15]. Beyond approximately N=15, validation using actual traffic count data begins to become infeasible due to TFM actions. As described in Section 2.1, observations where the deterministic trajectory modeler predicted a peak count greater than (MAP-6) were excluded from analysis. ADM2 and CRCT often agree on their peak count prediction,[12] thus when the ADM N is greater than (for example) 15, the CRCT peak predictions should also be in that vicinity. When the CRCT peak prediction is 15, the (MAP-6) rule excludes the observation for any sector whose MAP is less than 21, which equates to excluding 740 of the 782 NAS sectors defined at the time the data were collected—almost 95% of sectors. When the peak prediction is 16, the rule excludes any sector with a MAP less than 22—all but 2 of the sectors. Therefore, as N increases, the data included in the validation set from the training data become less and less a representative sample of NAS sectors.

The correlation results for N in [11, 15] are now presented. As with the other analyses reported in this section, the correlation was calculated between the probabilities according to the Gaussian model and those observed in the test dataset. For this analysis, as with the analyses for N in [5, 10], Altitude-LAT-N-Error combinations with less than 30 cases in the test set were excluded. There were 1924 combinations available for analysis. The R^2 for the correlation analysis conducted on this dataset was 0.9333. The scatter plot is shown in Figure 2-20.

[12] Although the ADM2 N's are of course not equivalent to the CRCT trajectory modeler's peak predictions, a moderate degree of correspondence is observed between them. This is not surprising, because both models predict traffic counts based primarily on the same input data—the schedules and flight plans of flights operating or expected to be operating in the NAS—though ADM2 uses a more complex set of weights and factors. See Section 2.5.2 for further discussion of the degree of this correspondence and its implications.

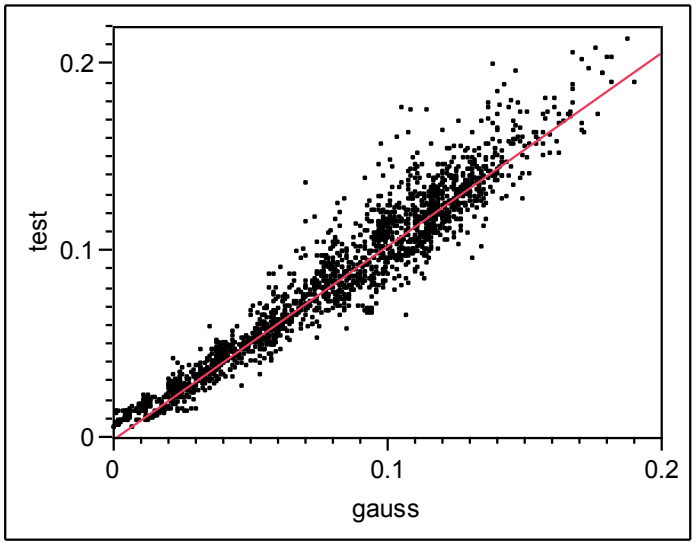

Figure 2-20. Scatter Plot of Gaussian Model and Test Data Error Probabilities for Each Combination of Altitude, LAT, N, and Error, for N in [11, 15]

The R^2 were also computed separately for each individual value of N. The purpose of this analysis was to determine where the cutoff should be placed, below which the training data PMF values would be used as the probabilities in the final model, and at or above which the PMFs derived from extrapolating the Gaussian parameters would be used. Even though the N range of [5, 10] was used for the Gaussian extrapolation, it was not required that the same cutoff be used for determining which N's would use actual (as opposed to modeled) probabilities in the final model. The goal was of course to use the probabilities which corresponded best to the actual test set probabilities for each N.

The R^2 for each subset of individual N are shown in Table 2-1, indicating how well the probabilities from the Gaussian model and from the observed training data each predicted the test data probabilities.

Table 2-1. R^2 of Linear Correlation between (Model and Test Data) and (Training Data and Test Data) Error Probabilities for Each Combination of Altitude, LAT, N, and Error, Separate for Each N

N	R^2 Gauss Model	R^2 Training Data	Number of Combinations
11	0.9750	0.9836	704
12	0.9334	0.9450	585
13	0.8781	0.8465	396
14	0.8314	0.8053	168
15	0.4409	0.2674	71

Based on these results, it was concluded that the training data histogram should be used in the final model for N's up to 12, with the probabilities computed from the extrapolation of the Gaussian parameters for N's of 13 and above.

2.5 Additional Analysis of Peak Count Prediction

As noted in [Wanke et al., 2008b], two types of analysis were required in order to complete the ADM2 construction and validation. In addition to extrapolating the error distributions to higher N's, (Sections 2.2 through 2.4), it was also necessary to conduct additional investigation of ADM2's performance at predicting peak count for various subsets of the data. This section describes the latter activity, comparing the deterministic (CRCT) model and ADM2 for various data subsets including time of day, day of week, and level of disagreement between the two models.

2.5.1 Time of Day/Day of Week

In [Wanke et al., 2008b], it was noted that we needed to analyze the extent to which model performance differed at various times of day or days of the week. The prediction errors were compared for the deterministic model and ADM2 for each hour of the day—the Coordinated Universal Time (UTC) hour[13] *for which* the prediction was made, not the *time when* the prediction was made. For signed error, negative values represent under prediction and positive represent over prediction. Figure 2-21 and Figure 2-22 show the mean signed and absolute error, respectively, by time of day.

[13] UTC is 4 hours ahead of Eastern Daylight Time (EDT), and thus is 5 hours from Central, 6 hours from Mountain, and 7 from Pacific. Therefore, for example, 0800 UTC is 4 AM EDT, 3 AM Eastern Standard Time, etc. Though a single value of UTC corresponds to several different local times in the contiguous U.S., the UTC analysis still provides a general picture of how the models perform at various times of day, especially during peak daylight hours versus midnight hours.

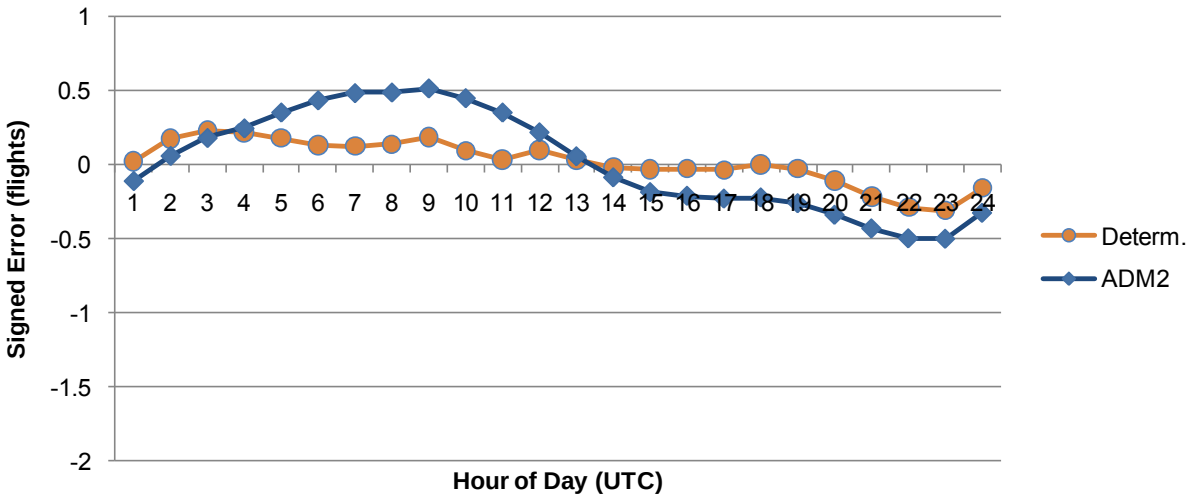

Figure 2-21. Mean Signed Error of Deterministic Model and ADM2 at Each UTC Hour of the Day

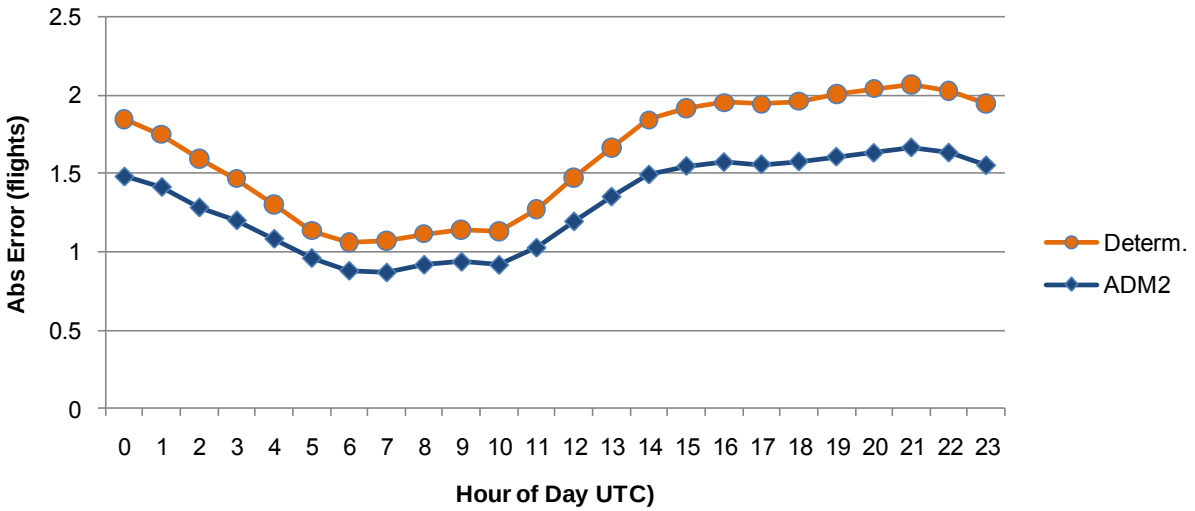

Figure 2-22. Mean Absolute Error of Deterministic Model and ADM2 at Each UTC Hour of the Day

These results show that ADM2 predictions have a tendency to slightly (by 0.5 flights or less) over predict the peak count in the middle of the night, and to under predict slightly during the busiest hours of the day. The CRCT model exhibits similar effects, though at a smaller magnitude. However the absolute prediction performance of ADM2 exceeds that of the deterministic model (lower absolute error) at all times of day. A signed error closer to zero does not necessarily mean that the average absolute performance will be higher, because near-zero mean signed errors may result from a combination of larger over predictions and larger under

predictions. A near-zero average signed error simply means little or no bias exists toward under or over prediction for the data subset.

ADM's slight tendency to over predict peak counts in the middle of the night is most likely due to the intercept term in the regression equation for ADM2. As described in detail in [Wanke et al., 2008a] and briefly at the beginning of Section 2, ADM2 adds a LAT-based factor to the known demand, essentially a "pop-up" factor. The value of this factor is based only on LAT and not designed to vary with time of day. Since the midnight hours have fewer pop-ups and fewer flights in general, the factor is sometimes not appropriate to apply at these hours, and an over prediction results. However, this effect, as well as the very slight under prediction during peak traffic hours, is small. Since at all times of day, the absolute error is still lower than that of CRCT—meaning that more often than not, the ADM modeling approach including the pop-up factor makes the model more accurate—no further adjustment to the model is considered necessary at this time.

Regarding day of the week, the comparison of ADM and the deterministic model for each individual day[14] also shows little differential performance for various days of the week or overall differences between the models. The mean signed and absolute errors are depicted in Figure 2-23 and Figure 2-24.

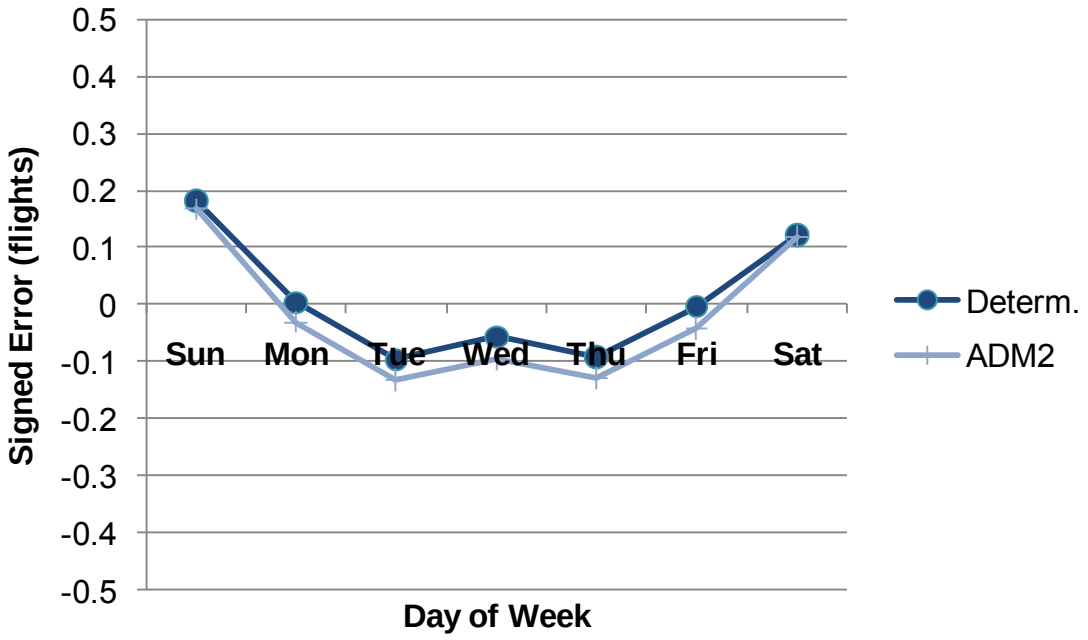

Figure 2-23. Mean Signed Error of Deterministic Model and ADM2 for Each Day of the Week

[14] The test data had between 3 and 5 different days' worth of data for each day of the week; details can be found in [Wanke et al., 2008a].

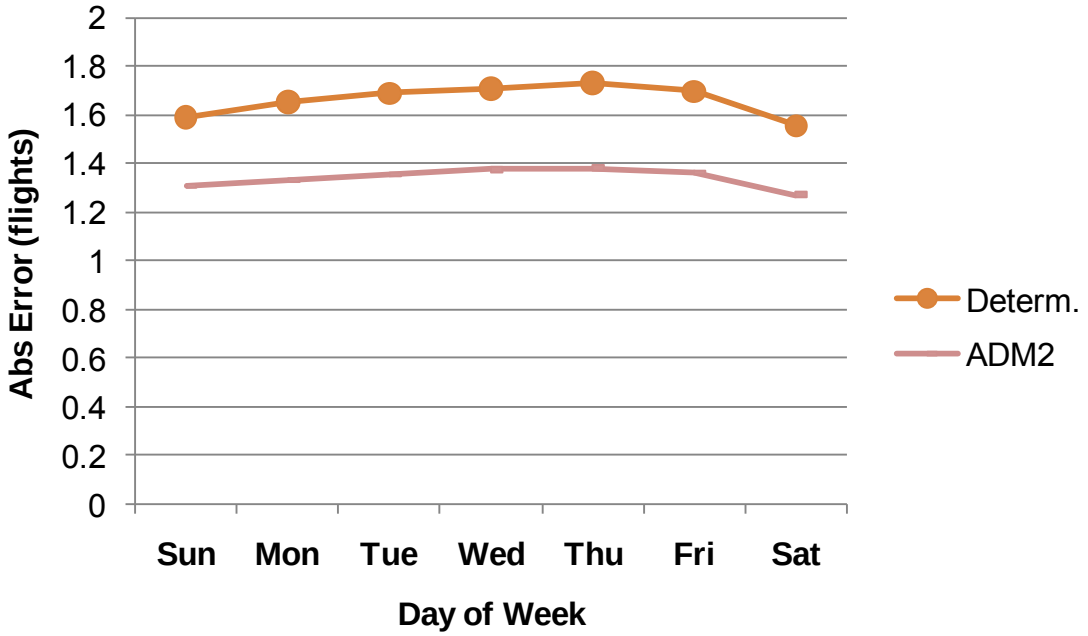

Figure 2-24. Mean Absolute Error of Deterministic Model and ADM2 for Each Day of the Week

It is interesting that both models tend to very slightly over predict demand on weekends and under predict midweek. This is believed to be because cancellations are more common on weekends, and pop-ups more likely in the middle of the week. The intercept factor in the ADM2 regression equations does account for an aggregation of pop-ups and cancellations. However, as with time of day, the model does not specifically account for day of the week when applying these factors. Because the day of week effect is so small, and since for all days of the week, ADM2's absolute error is lower than that of CRCT, further adjustment to the model is not considered necessary.

The comparisons in this section between time-based subsets of the data, like many direct aggregated comparisons of ADM and the deterministic model, do not make the ADM seem more accurate in an operationally meaningful way. Section 2.5.2, however, shows specific cases where the difference is more meaningful.

2.5.2 Disagreement between ADM and CRCT

A large part of the reason that ADM and CRCT seem to have roughly equivalent prediction performance when the data are aggregated across too many dissimilar cases, is that fairly often, the peak predictions of the two models are roughly equivalent. This can be seen from Figure 2-25, representing the PMF of the relative difference between the integer peak predictions of ADM2 and CRCT across all observations in the 28-day test dataset.

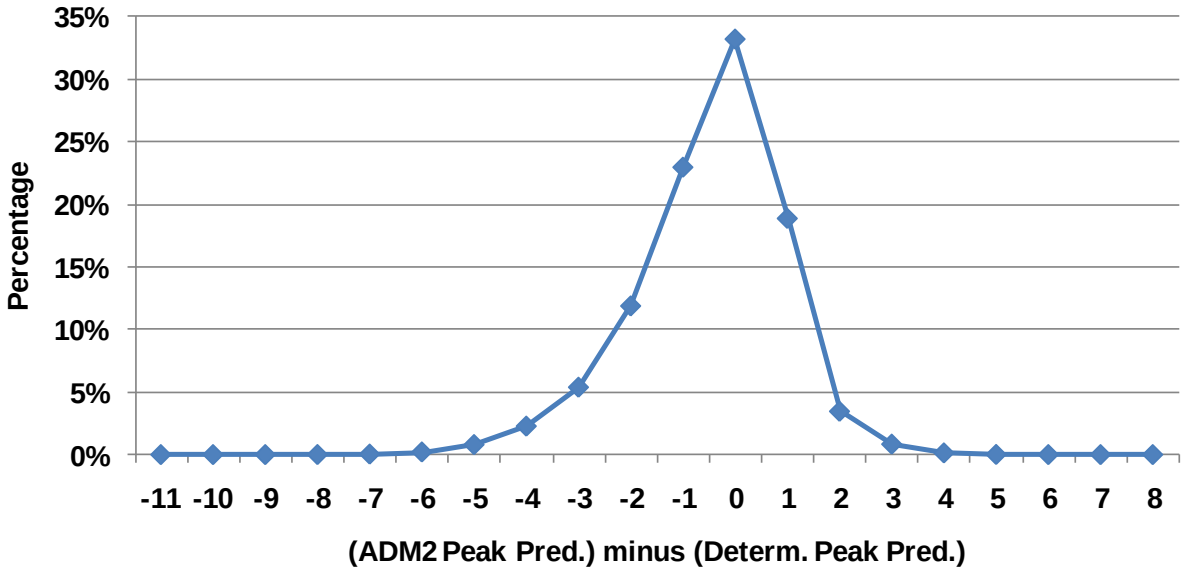

Figure 2-25. PMF of Difference between ADM2 and Deterministic Model's Peak Count Estimates, Across All Data

This distribution indicates that the two models agree on the peak prediction (horizontal axis=0) approximately 33% of the time, but that there is often disagreement as well, Naturally, if they seldom disagreed, there would be little point to comparing their prediction performance.

Figure 2-26 shows the relative performance of the two models in terms of absolute error for each value of difference between the models' peak predictions. Since difference values of -11, -10, 7, and 8 each had less than 30 observations in the entire dataset, they were excluded from analysis. Note that the horizontal axis in the figure, "(ADM2 Peak Pred.) minus (Determ. Peak Pred.)", is not to be confused with Error value. A value on the horizontal axis of -2, for example, means that the ADM2 prediction was 2 flights lower than the CRCT prediction, regardless of which model (if either) was more accurate. Correspondingly, a value of 2 means the ADM2 prediction was 2 flights higher, and so on.

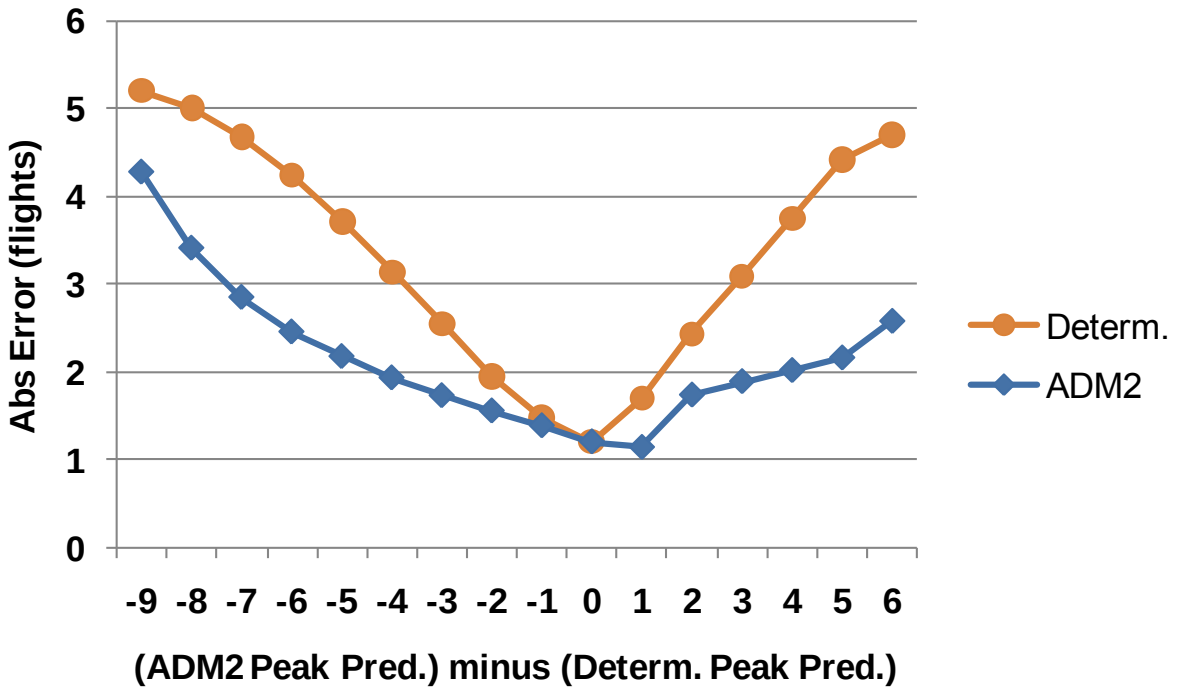

Figure 2-26. Mean Absolute Error of Deterministic Model and ADM2 at Each Level of Disagreement between the Two Models' Peak Count Predictions

This result indicates that the ADM2 prediction is, on average, more accurate than the deterministic model by several flights in those cases where it *can* be, that is, when there is enough disagreement between the models for a difference to be noticeable.

The obvious question that arises is under what circumstances the two models are most likely to disagree. A number of exploratory analyses were conducted in an attempt to characterize the situations where disagreement was most likely, and to assess whether some commonality could be found in the cases where ADM2 was most likely to outperform the deterministic model or vice versa. However, no consistent pattern emerged. Of course, in general, discrepancies between the models' predictions are more likely at high predicted and/or actual demand because there are more flights that could be incorrectly predicted by one of the models. However, no large performance differences were observed at any particular demand level, or at any other specific subset of data (e.g., specific LATs or sector types).

Therefore, it was concluded that the model did not require special adjustment for any particular subset of the data such as LAT or sector type. Rather, for almost any subset, the average performance of ADM2 was slightly more accurate than the deterministic model. Furthermore, though large disagreements are not specifically associated with any data subset, ADM2 does tend to predict peak counts several flights closer to the actual demand when disagreement does occur.

2.6 Summary and Next Steps

The ADM2 model created in 2008 [Wanke et al., 2008a] was completed by extending it to high-demand cases. Testing indicates that the ADM2 model provides moderately better peak count predictions than the deterministic CRCT model, and that this advantage is pronounced in cases where ADM2 and CRCT predictions disagree by more than 1 or 2 flights. It was also found that special adjustments for LAT, sector type, time of day, or time of week were not required; the ADM2 model was more accurate than the CRCT model for any subset of these.

The extension of the prediction error distributions to high-demand cases represents the completion of the last unfinished piece of the revised aggregate demand model. The completed model will be used in further work (Sections 5 and 6). To the extent possible, the extrapolated distributions were compared to the same test dataset used in [Wanke et al., 2008a]. Based on this analysis, it was concluded that the distributions generated from the extrapolation method predicted the test data very well. A comparison of the prediction accuracy results for the probabilities at various N led to the decision to implement the ADM2 distributions using the actual observed probabilities for $N \leq 12$, and the probabilities from Gaussian curve fitting and mean/SD extrapolation for $N > 12$. The high likelihood of the test data being skewed by demand-reducing ATM actions for N values greater than about 15 in the majority of NAS sectors rendered proper validation using the same method impossible for situations where N exceeded 15. Individual flight modeling using an extension of the Monte Carlo simulation method employed in the analyses described in Section 5 is recommended to validate the model for cases with very high demand. However, for the present, because of the excellent curve fits observed for the portion of the data that could be validated using this method, we have a high degree of confidence that the probabilities are reliable.

The ADM2 model has been implemented in CAASD's TFM analysis software toolkit, to support further testing and improvement of the model and to support the ongoing probabilistic demand and risk management research described elsewhere in this document.

Now that the model is complete, we plan to coordinate with the FAA regarding technology transfer to TFMS. For real-time application, we need to recommend reproducible methods for updating the model parameters to reflect changes in the NAS over time.

Further validation with the Monte Carlo simulation may help identify the key sources of prediction error in the model such that it can be improved. Also, if proposed operational capabilities (e.g., surface data reporting improvements) lead to reduced prediction error components (e.g., reduced departure time prediction error), this can be simulated to estimate the improvement in traffic demand prediction.

3 Capacity Prediction: Weather-Impacted Sector Capacity

Airspace capacity prediction uncertainty is more difficult to understand than traffic prediction uncertainty. Even in clear weather, there is not a fully-accepted definition of sector capacity. In convective weather conditions, even when the weather is perfectly forecasted, it is a complex problem to estimate the loss of capacity due to weather impact. Finally, weather forecasts are not only uncertain, but good mathematical models of weather forecast error are not generally available. Nonetheless, it is essential to address these problems in order to develop effective congestion resolution capabilities that work under convective weather conditions.

This section describes the continued development and validation of three models aimed at identifying and predicting the capacity of a sector which is impacted by convective weather. We compared the effectiveness of the three methods in predicting sector throughput reduction for different sector types and under different weather conditions, and we estimated the predictability of the weather impact metrics.

3.1 Weather Impact Indexes

The three models of weather impact on sector capacity explained in [Wanke et al., 2008a] introduced three sector weather impact indexes: two-dimensional (2-D) weather coverage, three-dimensional (3-D) weather avoidance altitude field (WAAF) coverage, and the flow-based reduced sector capacity ratio. Each of these will be described briefly in this section.

To calculate the weather impact indexes, we used the Corridor Integrated Weather System (CIWS). CIWS is a MIT/LL weather product, which provides accurate and high update rate information on storm locations and echo tops with one-kilometer (km) spatial resolution [Weber et al., 2006]. The precipitation intensity of the severe weather is characterized by the Vertically Integrated Liquid (VIL) metric in CIWS. Weather areas with measured VIL at and above level three (VIL3+) correspond to heavy and extreme precipitation in current FAA terminology.

3.1.1 2-D Weather Coverage

VIL3+ coverage in a sector is an important indicator of weather impact on sector capacity [Davis et al., 2005]. Without the help of automation, traffic managers often estimate the VIL3+ coverage in a sector by themselves and predict the reduction of the sector capacity based on their own experience. Thus, the first weather impact index examined, the 2-D weather coverage, is the percent of the sector area with VIL3+ coverage.

3.1.2 3-D WAAF Coverage

The 2-D weather coverage does not take into account the storm height, as measured by radar echo tops. Echo tops will likely be a major factor for high and super high altitude sectors, where flights may be able to fly over certain storms. Another key element to factor in is pilot avoidance behavior near severe weather. Flight in the vicinity of a severe weather area may be acceptable to some pilots but not to others. This acceptance is likely a function of many factors such as operator experience and training, risk aversion or acceptance, airline policy, the existence of

alternate options, and the expected amount of time that will be spent in the severe weather hazard space. Research is taking place on the behavior of pilots near severe weather. MIT/LL has developed their first Convective Weather Avoidance Model (CWAM1), which models the pilot deviation behavior in and around severe weather as a function of explanatory variables (such as reflectivity level and echo tops), by observing actual flight tracks around severe weather cells [DeLaura and Evans, 2006]. Note that CWAM1 is an initial model; MIT/LL is developing CWAM2 and CWAM3 [DeLaura et al., 2008].

With echo tops and the pilot deviation behavior model, the weather area in a sector that most flights would avoid can be identified. The deviation decision model in [DeLaura and Evans, 2006] shows that most flights fly deltaZ above the 90th percentile of the echo top height in the 16X16 km^2 neighborhood, where deltaZ is a function of the VIL3+ weather coverage in the 60X60 km^2 neighborhood. There are two versions of this function, deterministic and probabilistic. With deterministic CWAM1, the deltaZ is treated as certain given the VIL3+ weather coverage in the 60X60 km^2 neighborhood; the weather avoidance altitude of each pixel—1x1 km^2, the resolution of CIWS—in a sector is then the deltaZ plus the 90th percentile of the echo top height in the 16X16 km^2 neighborhood [Song et al., 2007b].

With probabilistic CWAM1, the deltaZ is a function of the probability of deviation. For example, if the VIL3+ weather coverage in the 60X60 km^2 neighborhood of the cell is between 70 and 80 percent, and the difference between the flight altitude and the 90th percentile of the echo top height in the 16X16 km^2 neighborhood of the cell ranges from -2,000 to +2,000 feet, then 80 percent of pilots would deviate around the cell. The weather avoidance altitude of each pixel can be calculated as follows:

1. Calculate the percentage of VIL3+ pixels in the 60 km neighborhood around the pixel (e.g., 75%).
2. Calculate the 90th percentile of the echo top height from the 16 km neighborhood of the pixel (e.g., 32,000 feet).
3. Determine the desired threshold for the probability of deviation (e.g., 0.8),
4. Find the range of deltaZ (in thousands of feet) from the probabilistic CWAM1 (Table 3-1) by going to the column corresponding to the percentage of VIL3+ pixels calculated in Step 1 and finding the probabilistic threshold from Step 3 in that column. Then the row or rows that contain that value correspond to the deltaZ range(s) that apply (e.g., for VIL3+ coverage of 75% and probability threshold of 0.8, deltaZ is between -2,000 and +2,000 feet).
5. Add the deltaZ range and the 90th percentile of the echo top height calculated in Step 2 to get the weather avoidance altitude range of the pixel (e.g., given VIL3+ coverage of 75%, probability threshold of 0.8, and 90^{th} percentile of 32,000 feet, the weather avoidance altitude is between 30,000 and 34,000 feet).

Table 3-1. Probabilistic CWAM1 Lookup,
Based on Personal Communication with MIT/LL CWAM1 Development Team

%VIL3+ / deltaZ	0 – 10%	10 – 20%	20 – 30%	30 – 40%	40 – 50%	50 – 60%	60 – 70%	70 – 80%	80 – 90%	90 – 100%
< -10	0.9	0.9	0.9	0.9	0.9	1	1	1	1	1
-10 to -6	0.8	0.8	0.8	0.9	0.9	1	1	1	1	1
-6 to -2	0.6	0.7	0.7	0.8	0.8	0.9	0.9	1	1	1
-2 to +2	0.4	0.5	0.6	0.6	0.7	0.8	0.8	0.8	0.8	0.8
+2 to +6	0.2	0.3	0.4	0.4	0.5	0.6	0.6	0.6	0.6	0.6
+6 to +10	0.1	0.2	0.2	0.3	0.3	0.4	0.4	0.4	0.4	0.4
> +10	0	0.1	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.3

To compare the effect of different deltaZ on sector capacity, the lower end of deltaZ, which is most aggressive and gives the highest capacity, and the higher end of deltaZ, which is most conservative and gives the lowest capacity, are used from Step 3 to calculate the weather avoidance field of the pixel in Step 4. For the example above, the aggressive value is 30,000 feet and the conservative value is 34,000 feet. In addition, to compare the effect of different probability thresholds on sector capacity, two different probability thresholds are selected (0.6 and 0.8) to calculate the weather avoidance field of each pixel. Thus, five types of WAAF are generated based on the deterministic and probabilistic CWAM1:

- Deterministic WAAF (referred to as WAAF-D)
- Aggressive WAAF, probability threshold 0.6 (WAAF-A6)
- Aggressive WAAF, probability threshold 0.8 (WAAF-A8)
- Conservative WAAF, probability threshold 0.6 (WAAF-C6)
- Conservative WAAF, probability threshold 0.8 (WAAF-C8)

The WAAF also indicates that the avoided weather size, shape, and location vary with altitude. Since aircraft fly at different flight levels in the sector, it is important to understand how the weather would impact each flight level of the sector. Thus, a sector is sliced into 1,000-foot altitude bands. For example, Sector ZID66 covers an altitude range of 23,000 to 33,000 feet, so ZID66 will have 10 altitude bands, from band 23 to band 32. A pixel in band X needs to be avoided if the weather avoidance altitude of that pixel is greater than X. The percentage of weather avoidance area for each sector altitude band can then be calculated by dividing the number of avoided pixels by the total number of pixels in each sector altitude band. The percentage of sector WAAF coverage is then the weighted sum of the percentages of weather avoidance area for each sector altitude band, where the weights reflect the observed usage of flight levels in the sector.

$$\begin{aligned} &Sector\ \text{WAAF}\ Coverage \\ &= \sum_i w_i \times BandCoverage(i) \\ &w_i : Weight\ on\ Altitude\ Band\ i \end{aligned} \qquad \text{Equation 3-1}$$

The altitude weights are important. If we assume that altitude usage is uniform in sectors, the impact of a storm with 30,000 foot echo tops in a sector that handles flights from 24,000 to 35,000 feet will be overestimated. The majority of the flights in that sector will be free of the weather near the top of the sector. Altitude usage is also important in deciding what to do about very high altitudes, such as altitudes above 40,000 feet. In many cases these altitudes will be above the weather impacts, but assuming that flights in the sector will be using these altitudes with the same frequency as other altitudes would be incorrect, since very few aircraft can operate at these altitudes. Figure 3-1 shows an altitude usage profile over the entire NAS for the month of June 2007. Here the usage is measured in total flight-minutes at each altitude.

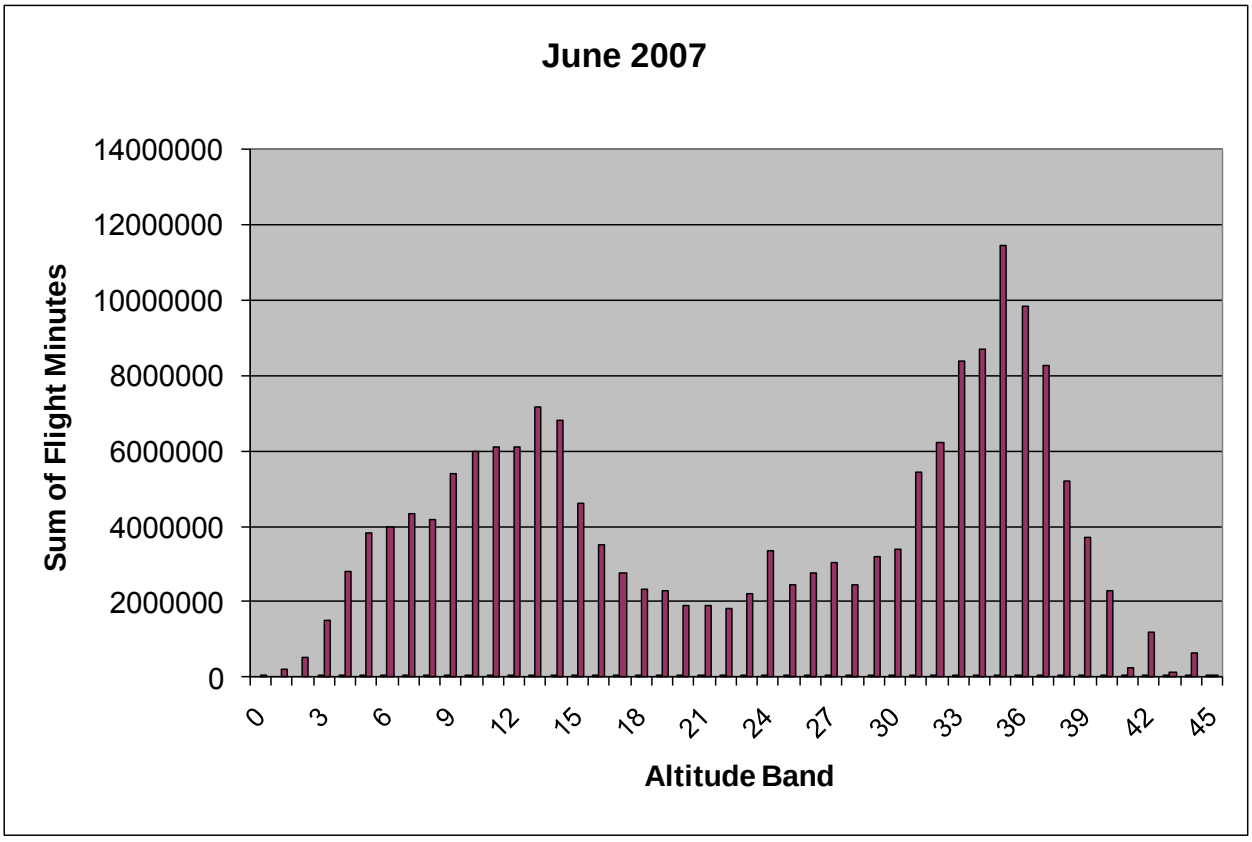

Figure 3-1. NAS Altitude Usage Profile

The distribution of altitude usage is a reflection of flight operators seeking the most efficient altitude for flight operations. We have assumed that this distribution will hold in most en-route airspace. If so, a single distribution can be used for most sectors. Some sectors, for example those handling arrival and departure flows, may not match this altitude usage distribution well, and may require special handling with sector-specific profiles. In this analysis, the altitude profile shown in Figure 3-1 was used for all NAS sectors. To evaluate the benefit of weighting the altitude band coverage with the NAS-wide altitude profile, the equally-weighted 3-D WAAF coverage is also calculated with each altitude band treated as equal (that is, assigned the same value of w_i in Equation 3-1).

3.1.3 Flow-Based Reduced Sector Capacity Ratio

In addition to the size of the weather area—more accurately, the WAAF area—the sector capacity is also highly correlated with the shape and location of the weather or WAAF area. Small storms located at critical locations can have more impact than larger storms in less critical locations. Figure 3-2 shows an example sector impacted by weather of the same shape and size, but at different locations in Case A and Case B. In Case A, the storm is located at a critical point, blocking major flows that account for most of the sector traffic. In Case B, only minor flows are blocked. The sector capacity reduction should be different for Case A and Case B. Thus, the third weather impact index captures the flow and flow pattern impact in a sector [Song et al., 2007a], where flows in a sector are defined to be the sector transit triplets (entry sector – current sector – exit sector).

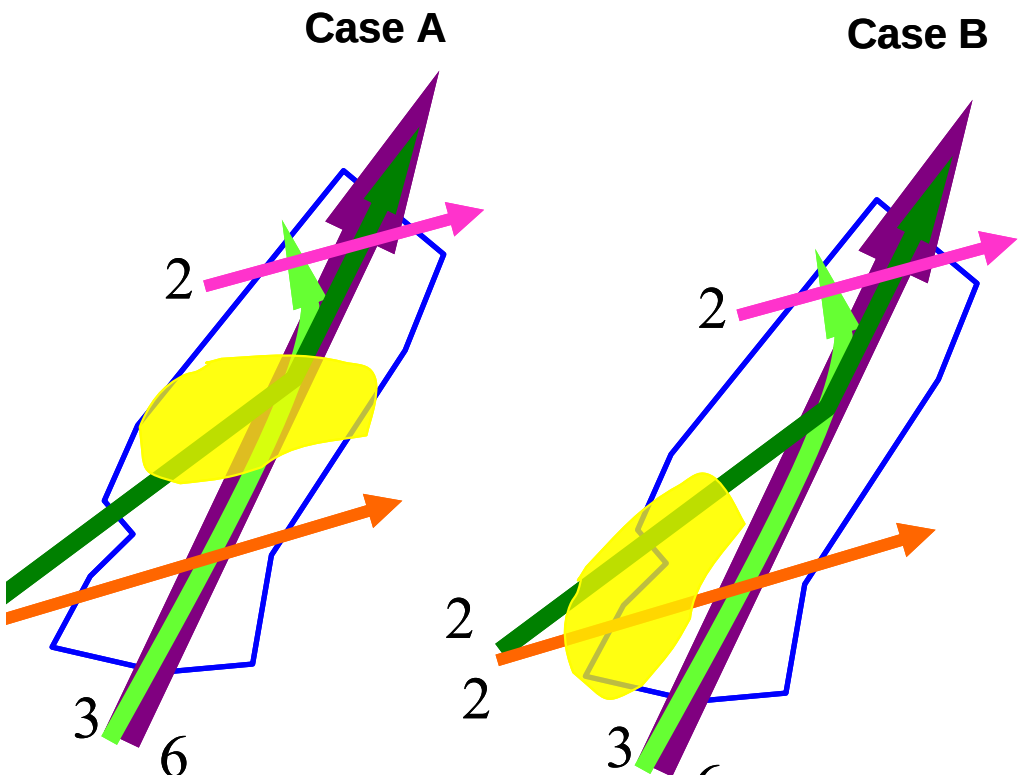

Figure 3-2. Example Sectors Under Severe Weather Impact

Rather than measure the flow (or triplet) blockage directly, its converse, the ratio of available flow capacity, is calculated with the minimal-cut (mincut) of the flow given the weather avoidance altitude field in the sector [Song et al., 2007b; Krozel et al., 2007]. For example, one flow for Sector B is A-B-C (Figure 3-3), which shows the flow is from Sector A, through B, and into Sector C for one altitude band. Each altitude band of a sector is a polygon. The yellow blocks in the figure are the WAAF area in that sector altitude band. Based on the generalized max-flow min-cut theorem, the capacity of flow A-B-C at each altitude band in Sector B is

dictated by its bottleneck, the mincut from the top edge T to the bottom edge B of Sector B avoiding the WAAF areas (Wmincut). The top edge T and the bottom edge B are the portions of the sector boundary clockwise and counterclockwise between the source edge S (the sector boundary shared by Sector A and Sector B) and the destination edge D (the sector boundary shared by Sector B and Sector C), respectively.

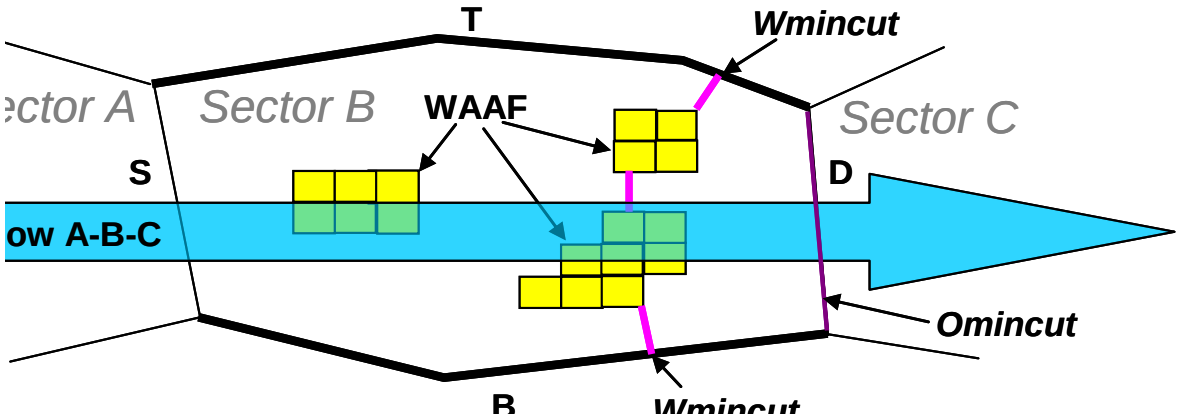

Figure 3-3. Flow Capacity Restricted with Mincut

Then for that particular flow, the available flow capacity ratio is the flow mincut ratio:

$$AvailableFlowCapacityRatio = \min_i \frac{Wmincut_i}{Omincut_i} \quad \text{Equation 3-2}$$

where $Wmincut_i$ is the mincut at altitude band i with WAAF area and $Omincut_i$ is the mincut at altitude band i without WAAF area. The minimum ratio is taken over the altitude bands used by the flow.

For the sector, the available *sector* capacity ratio is the weighted average of the available *flow* capacity ratios of all the flows in the predicted traffic flow pattern:

$$AvailableSectorCapacityRatio = \sum_{j=1}^{m} W_j \times AvailableFlowCapacityRatio \quad \text{Equation 3-3}$$

where m is the total number of flows; and W_j is the weight on flow j, which is the number of flights on flow j divided by the total number of flights in the sector. Finally, the *reduced* sector capacity ratio is 1 minus the available sector capacity ratio.

3.2 Estimated Actual Sector Capacity

Sector capacity as an indicator of controllers' workload threshold is dependent on the complexity of the traffic flows within the sector, as well as the presence or absence of hazardous weather.

No one really knows what the actual sector capacity should be for each 15-minute period. The historical total number of flights the sector handled in a 15-minute period (sector throughput) can give us some knowledge of the sector capacity in the current operational environment. However, many variables that affect the sector capacity are difficult to isolate when collecting the historical sector throughput.

One research study [Davis et al., 2005] estimated the weather impacted capacity for a sector to be the upper bound of the sector throughput in all cases with the same weather coverage in the sector. The upper bound of the sector throughput could both underestimate and overestimate the sector capacity. What has been through the sector is not the same as what can be handled in the sector. In many cases there are not enough flights that are planned to fly through the sector at the time of the observation to achieve the maximum throughput. In other cases, flights are hampered from reaching the sector due to weather or congestion in other sectors. Also demand in a sector could be reduced due to ATM initiatives that have anticipated the weather. There can be other cases where the observed throughput is higher than what a weather-impacted sector can expect to handle. This can happen when special traffic patterns are implemented to move flights around weather. These patterns may limit merging and crossing traffic in critical sectors to allow higher throughput than would be possible with the typical traffic patterns. Also, there can be other cases where the controller workload was higher than acceptable, sometimes due to unexpected rapid weather development.

In our FY09 research, we collected all the observations of actual sector throughput in June and July 2007. Observations were filtered out when a sector had low predicted demand one hour before the observation. This filter attempts to eliminate cases where the ATM system is reducing traffic in anticipation of the weather. ATM initiatives often happen more than one hour before an event, and can be overly restrictive due to the uncertainty in forecasting weather impacts.

The remaining observations were then binned according to the sector weather impact indexes. The bin sizes were carefully chosen to reflect the varying ranges of the different indexes, so that sample sizes were similar across the indexes when computing correlations. Most of the observations fall between 0% and 20% for 2-D weather coverage and between 0% and 50% for 3-D WAAF coverage, while flow-based reduced sector capacity ratio observations range up to 100%. So, the filtered sector throughput observations were binned by every 2% of 2-D weather coverage, every 5% of 3-D WAAF coverage, and every 10% of flow-based reduced sector capacity ratio. Within each bin of sector throughput observations, the top two and bottom two data points were deleted as outliers. The high throughput outliers may represent cases where workload was unacceptably high or cases where special high throughput flow patterns were used. The estimate of the actual sector capacity for each weather coverage bin was calculated from the 95th percentile of the throughput values if there were more than five data points in the bin.

3.3 Linear Correlation between Estimated Actual Sector Capacity and Weather Impact Indexes

In last year's analysis, we used WAAF-D—based on deterministic CWAM1—to calculate the 3-D WAAF coverage and the flow-based reduced sector capacity ratio. This year, we compared the weather impact indexes for all forty-eight high sectors from four northeast air traffic control

centers (ARTCCs): New York (ZNY), Washington (ZDC), Indianapolis (ZID), and Cleveland (ZOB). For these sectors, the linear correlations for the following variables were examined:

- Between each of the three sector weather impact indexes (2-D weather coverage, 3-D WAAF coverage, and the flow-based reduced sector capacity ratio) and the estimated actual sector capacity
- Between the 95th percentile of the flow throughput (assumed to be the estimated actual flow capacity) and the available flow capacity ratio for the top three major flows of the sector
- Between the estimated actual sector capacity and the available flow capacity ratio for the major flow of the sector

The reduced sector capacity under severe weather impact is heavily dependent on the operational usage of the sector. In general, statistically-significant linear correlations were found between the estimated actual sector capacity and the three sector weather impact indexes. None of the three sector weather impact indexes has the strongest linear correlation with the estimated actual sector capacity for all the sectors examined.

The results for each of the 48 high sectors we examined are shown in Figure 3-4. The sectors are color coded to show the sector weather impact index that has the strongest linear correlation with the estimated actual sector capacity. The 11 blue sectors are the sectors where 2-D weather coverage has the strongest linear correlation with the estimated actual sector capacity. The 2 sectors where equally-weighted 3-D WAAF coverage has the strongest correlation are shown in red, and the 3 sectors where altitude-weighted 3-D WAAF coverage has the highest correlation are shown in green. The 32 purple sectors are where the flow-based reduced sector capacity ratio has the strongest linear correlation with the estimated actual sector capacity.

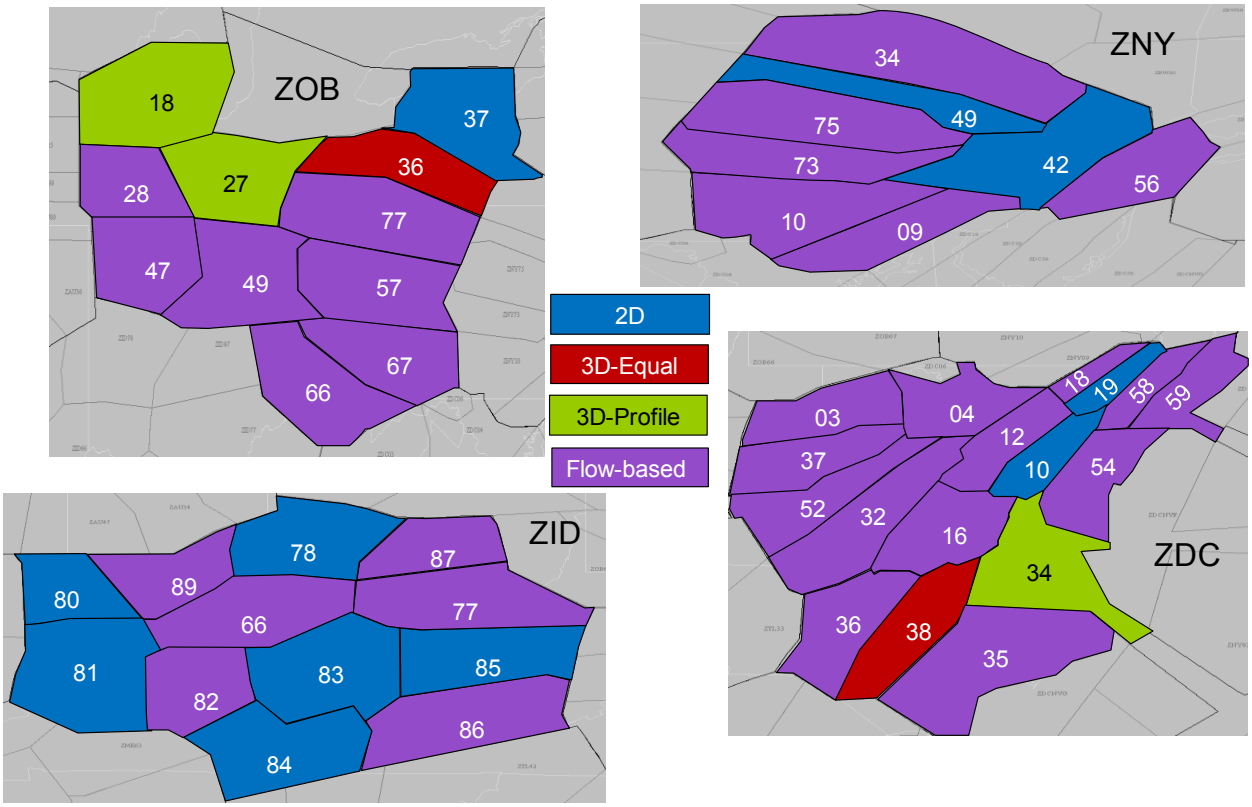

Figure 3-4. Sector Weather Impact Index with Strongest Correlation with Estimated Actual Sector Capacity for High Sectors

Generally, the flow-based reduced sector capacity ratio has the strongest correlation in sectors with dominant flows. Figure 3-5 shows the linear correlations of an example sector (ZDC12) with a dominant flow (ZDC16-ZDC12-ZDC18). The sector weather impact index for the blue line is the 2-D weather coverage, for the red line is the equally-weighted 3-D WAAF coverage, for the green line is the altitude profile weighted 3-D WAAF coverage, and for the purple line is the flow-based reduced sector capacity ratio.

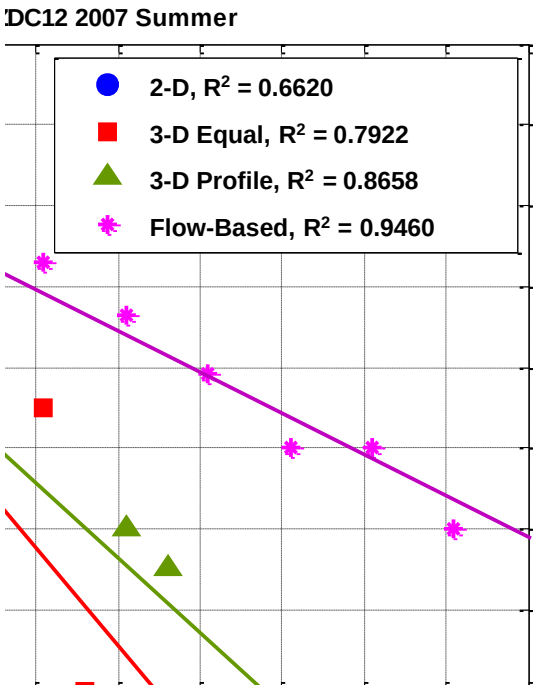

Figure 3-5. Sector Throughput versus Reduced Sector Capacity for ZDC12

Comparing the blue dots with the red squares in Figure 3-5, the correlation between the estimated actual sector capacity and the 3-D equally weighted WAAF coverage (0.7922) is stronger than the correlation for the 2-D weather coverage (0.6620). This shows that the deterministic CWAM1 works well in ZDC12 since WAAF is built upon the deterministic CWAM1. Comparing the green triangles with the red squares in Figure 3-5, the altitude profile added some value to the correlation for the 3-D WAAF coverage (0.8658 versus 0.7922). Comparing the purple stars with the rest in Figure 3-5, the correlation for the flow-based reduced sector capacity ratio is stronger than the correlations for all the other sector weather impact indexes (for example, 0.9460 for flow-based versus 0.6620 for 2-D).

For sectors with dominant flows, the 95th percentile of the dominant flow throughput has strong linear correlation with the *available* flow capacity ratio of the dominant flow, but not for the other flows. For example, consider ZDC12, a high sector that handles traffic arriving at New York and Philadelphia airports. Figure 3-6 shows the top three major flows (ranked by flight count) through ZDC12: ZDC16-ZDC12-ZDC18 (red), ZDC16-ZDC12-ZDC17 (blue), and ZDC72-ZDC12-ZDC18 (green). ZDC16-ZDC12-ZDC18 is the dominant flow of Sector ZDC12. During severe weather impact, the available flow capacity ratio for ZDC16-ZDC12-ZDC18 has strong linear correlation with the 95th percentile of the flow throughput, but the other two flows do not, as shown in Figure 3-7.

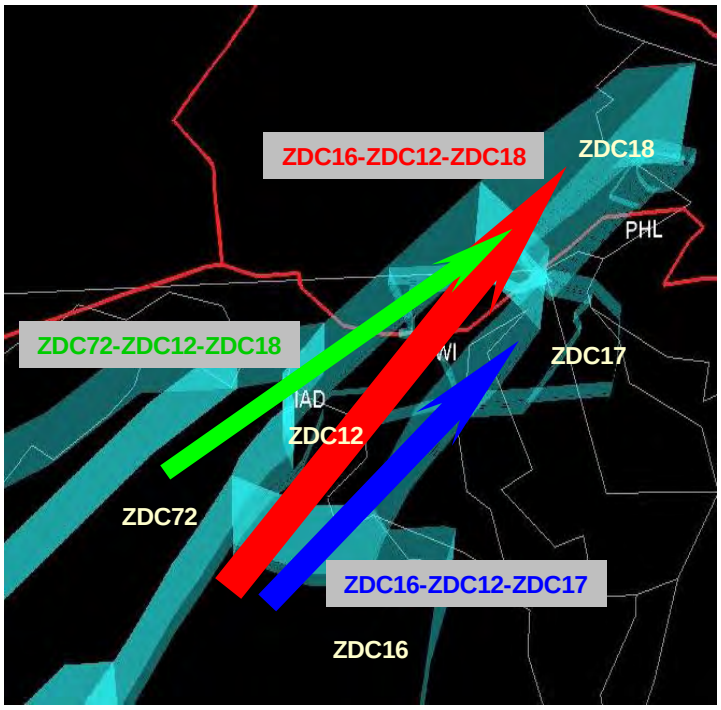

Figure 3-6. ZDC12 Major Flows

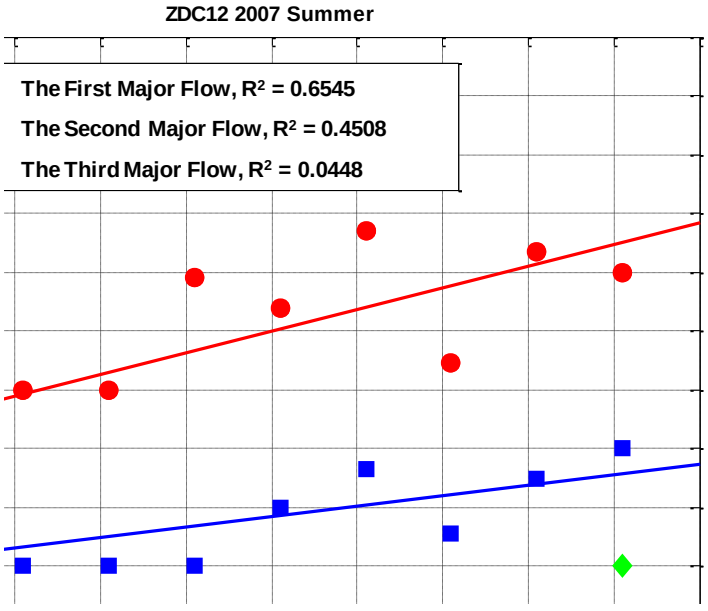

Figure 3-7. ZDC12 Top Three Flows Throughput

This result reflects the sector's current operational usage during severe weather impact. To reduce controller workload and improve sector efficiency during severe weather, ZDC12 generally continues to handle all traffic on the dominant flow while traffic on the other flows is reduced or eliminated. To verify this observation, some in-house former controllers and traffic flow managers were interviewed. Their explanation about how they handled traffic during severe weather matches the results shown in Figure 3-7 and Figure 3-8. Figure 3-8 shows the estimated actual sector capacity of ZDC12 as a function of the available flow capacity ratio of ZDC16-ZDC12-ZDC18. There is strong correlation between the available flow capacity ratio of ZDC16-ZDC12-ZDC18 and the estimated actual sector capacity (0.7106), even stronger than the correlation between the 2-D sector weather coverage and the estimated actual sector capacity (0.6620, blue line in Figure 3-5).

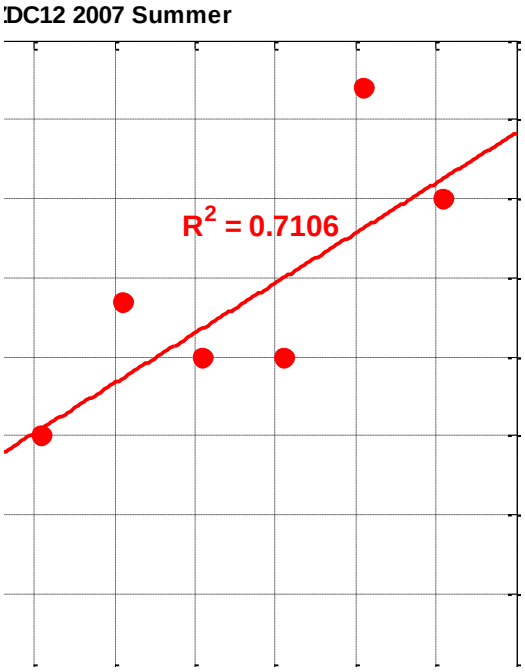

Figure 3-8. ZDC12 Sector Throughput versus Available Flow Capacity

For the sectors in which deterministic CWAM1 does not work well, the 2-D weather coverage has the strongest linear correlation with the estimated actual sector capacity (blue sectors in Figure 3-4) since both 3-D WAAF coverage and the flow-based reduced sector capacity ratio are based on deterministic CWAM1 in this analysis. For these sectors, we calculate the flow-based reduced sector capacity ratio again with four additional types of WAAF (WAAF-A6, WAAF-A8, WAAF-C6, and WAAF-C8 as defined in Section 3.1.2) based on probabilistic CWAM1. All the blue sectors in Figure 3-4 turn to purple with some types of WAAF calculated with probabilistic CWAM1. For example, as shown in Figure 3-9 for ZID85, the flow-based reduced

sector capacity ratio calculated with WAAF-A8—labeled Flow-A8—has the strongest linear correlation with estimated actual sector capacity.

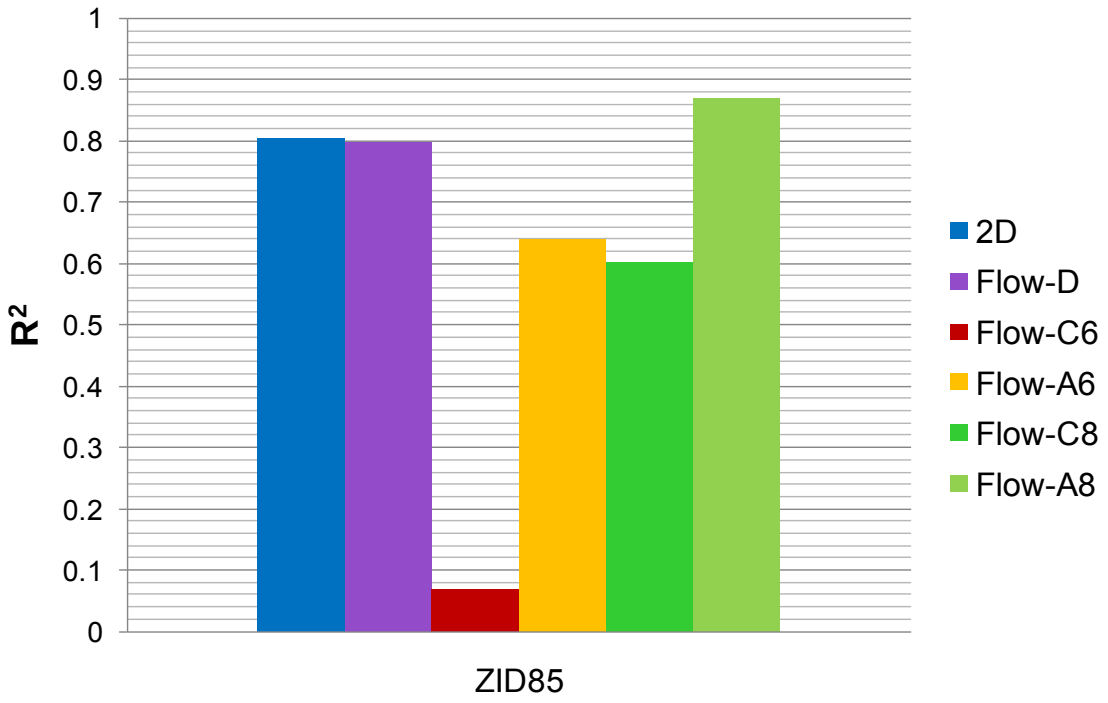

Figure 3-9. Comparison of Different Types of WAAF for ZID85

Remember that WAAF-A8 is the WAAF calculated with probabilistic CWAM1 when the probability threshold is set to 0.8 and deltaZ is aggressive (the lower end of its range). Flow-C6 in Figure 3-9 means the flow-based reduced sector capacity ratio calculated with WAAF-C6—setting the probability threshold to 0.6 and deltaZ to conservative (the higher end of probabilistic CWAM1). Flow-D is the flow-based reduced sector capacity ratio calculated with deterministic WAAF.

On the other hand, for sectors such as ZID83, the flow-based reduced sector capacity ratio calculated with WAAF-C8 has the strongest linear correlation with estimated actual sector capacity, as shown in Figure 3-10. For this sector, Flow-A6, Flow-A8 and Flow-C8 have better linear correlation with estimated actual sector capacity than does Flow-D, but Flow-C8 is significantly better than the others. Figure 3-11 shows the comparison of different types of WAAF for *all* the blue sectors in Figure 3-4.

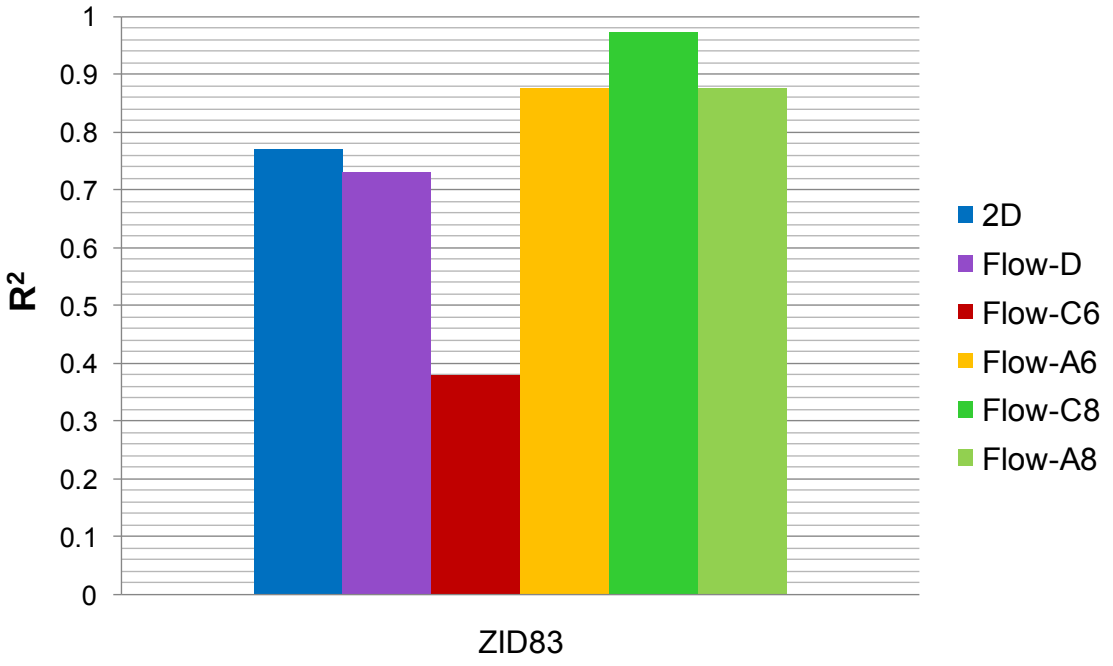

Figure 3-10. Comparison of Different Types of WAAF for ZID83

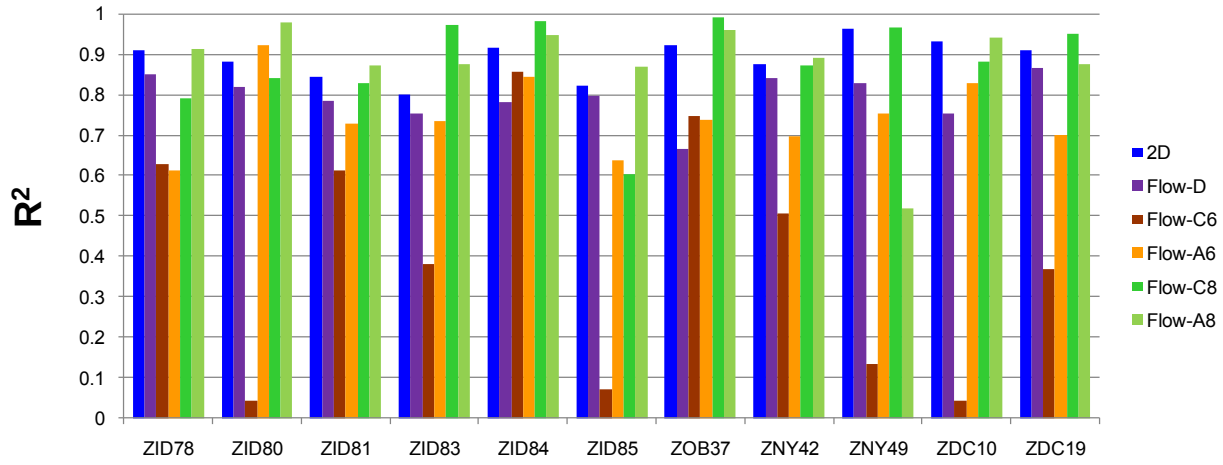

Figure 3-11. Comparison of Different Types of WAAF for All Blue Sectors in Figure 3-4

As already shown in Figure 3-4, the 3-D WAAF coverage has the strongest linear correlation with estimated actual sector capacity for some of the sectors (red and green sectors in Figure 3-4). These sectors have dominant flows that are transitioning (climbing or descending). When calculating the mincut for a transitioning flow, that flow is projected to the level flow that goes through the midpoint of the line connecting the average entry and exit points of the transitioning flow [Song et al., 2007b]. This treatment of the transitioning flows in the flow-

based model may be the reason that the 3-D WAAF coverage has the stronger linear correlation for sectors with dominant flows that are transitioning flows. Further analysis is necessary to check this hypothesis.

For sectors where the 3-D WAAF coverage has the strongest correlation with estimated actual sector capacity, sometimes the best approach was to use equally-weighted altitude bands (red sectors in Figure 3-4), and sometimes the NAS-wide altitude usage profile-weighted approach worked better (green sectors in Figure 3-4). The reason is that the NAS-wide altitude usage profile does not match the sector altitude usage profile very well for some sectors. For example, Figure 3-12 compares the altitude usage profile of ZOB36 and the NAS for the altitude band range of ZOB36. In this example, the extreme case is for altitude bands 32 and 33. Altitude bands 32 and 33 are the least frequently used in ZOB36, while they are the most frequently used over the NAS. Further research is necessary to determine whether using sector-specific altitude profiles would improve the correlations for the 3-D WAAF coverage index.

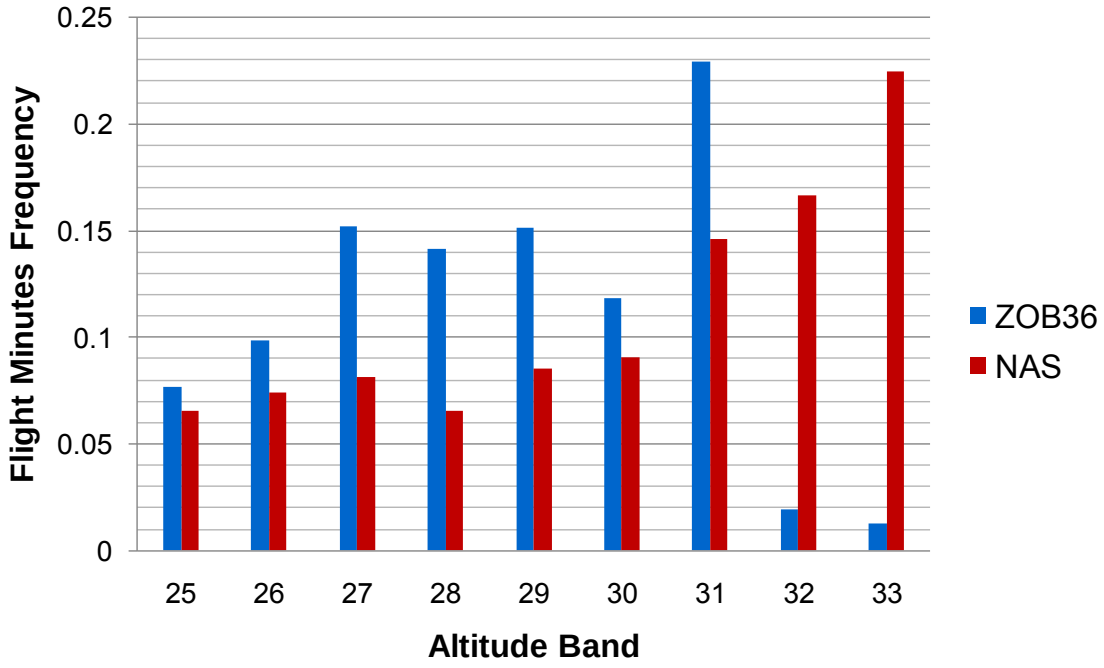

Figure 3-12. Altitude Profile Comparison between ZOB36 and NAS

Forty-three low sectors—all those with enough weather and traffic data for the analysis—were also selected from the same four northeast air traffic control centers (ZNY, ZDC, ZID, and ZOB) for comparing the weather impact indexes. Note that two of these sectors (ZNY75 and ZNY56) are both low and high sectors. The 3-D WAAF coverage and flow-based reduced sector capacity ratio were calculated with five types of WAAF (WAAF-D, WAAF-A6, WAAF-A8, WAAF-C6, and WAAF-C8 as defined in Section 3.1.2) based on both deterministic and probabilistic CWAM1. The linear correlations between the weather impact indexes and estimated actual sector capacity were examined and compared.

An example of the correlations, for Sector ZNY35, is shown in Figure 3-13. The leftmost (blue) bar is the R^2 between the 2-D weather coverage and estimated actual sector capacity. The next five (red) bars are for equally weighted 3-D WAAF coverage calculated with five types of WAAF (WAAF-D, WAAF-A6, WAAF-C6, WAAF-A8, WAAF-C8), so 3DEC6 means the equally weighted 3-D WAAF coverage calculated with WAAF-C6. The following five (green) bars are for altitude usage profile weighted 3-D WAAF coverage calculated with five types of WAAF. And the last five (purple) bars are for the flow-based available sector capacity ratio calculated with five types of WAAF. For ZNY35, the 2-D weather coverage has the strongest linear correlation with estimated actual sector capacity.

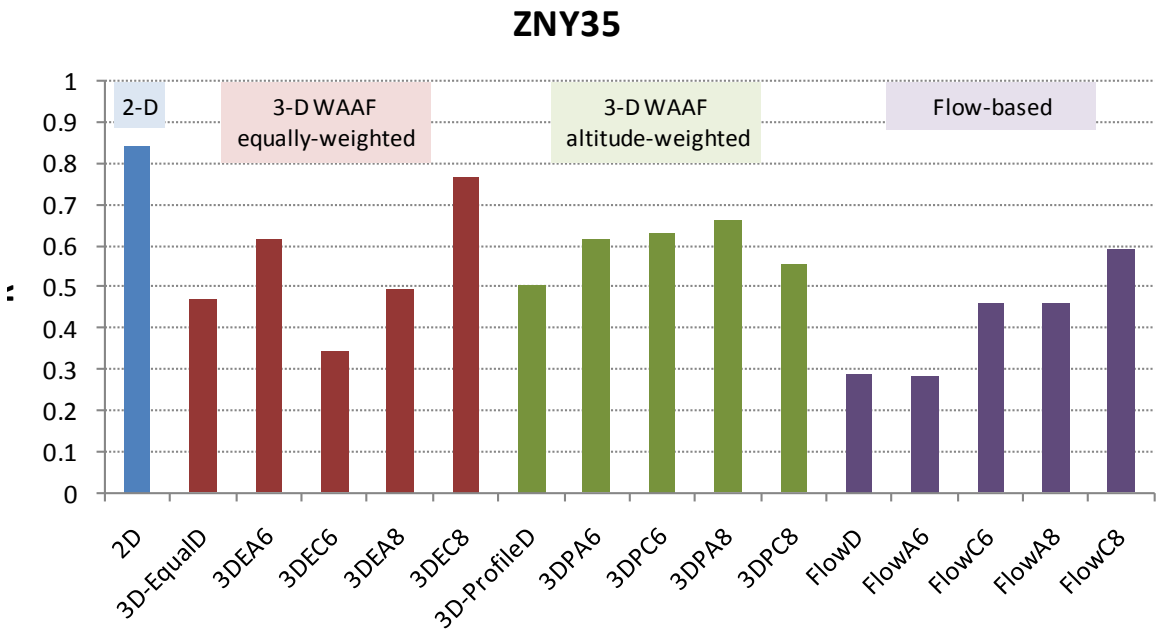

Figure 3-13. Comparison of Linear Correlations for ZNY35

The results for all 43 low sectors we examined are shown in Figure 3-14. The sectors are color coded to show the sector weather impact index that has the strongest linear correlation with estimated actual sector capacity. As we can see from Figure 3-14, there is no single dominant color for these 43 low sectors. Neither deterministic nor probabilistic CWAM1 works well for low sectors.

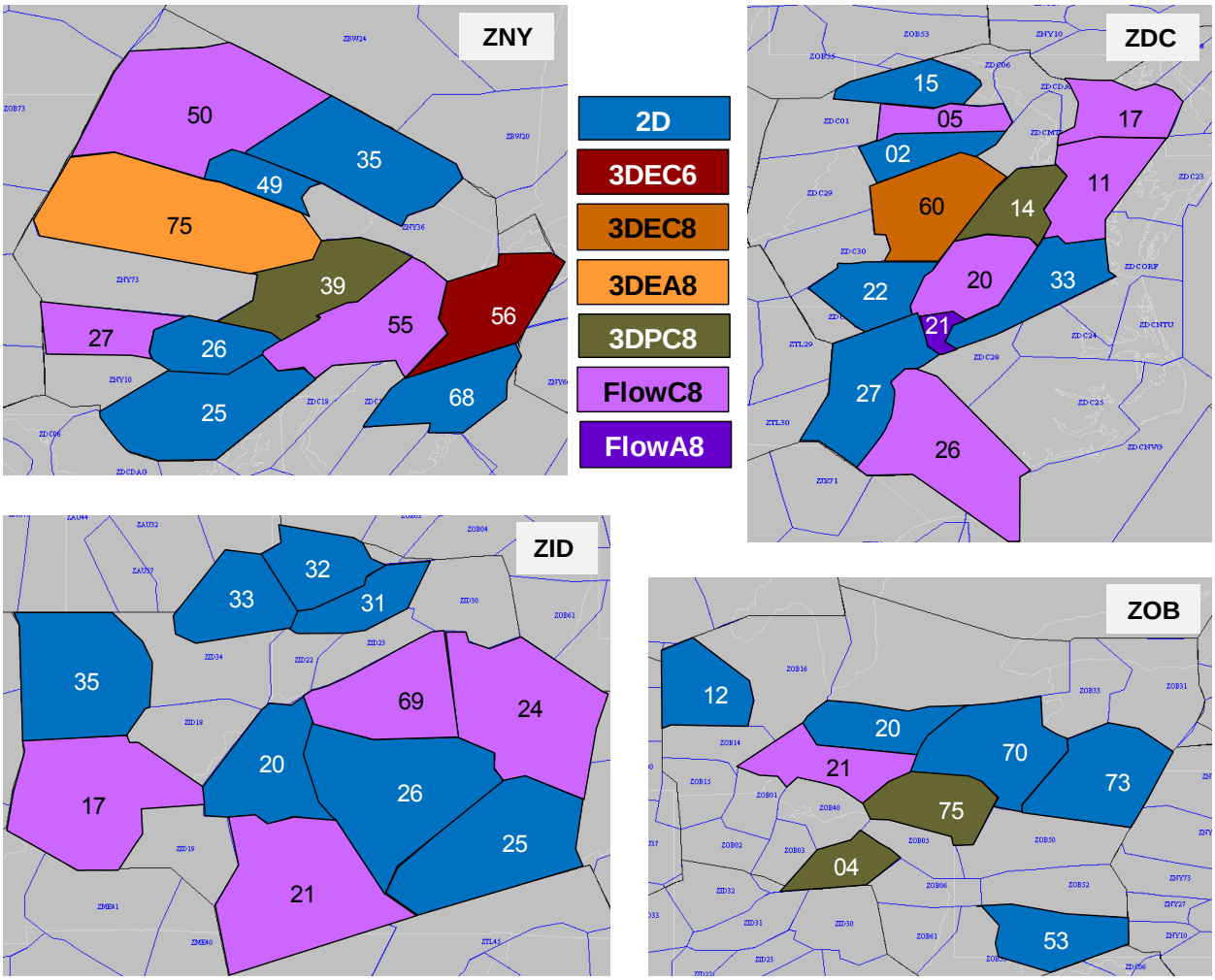

Figure 3-14. Sector Weather Impact Index with Strongest Correlation with Estimated Actual Sector Capacity for Low Sectors

3.4 Predictability of Sector Weather Impact Indexes

To be able to predict weather-impacted sector capacity, the predictability of the sector weather impact indexes has to be examined. To be able to examine just the weather impact, not how predictable the traffic demand is in a sector, we chose not to analyze the predictability of the available *sector* capacity ratio, which is weighted by the traffic demand on each flow. Instead, we analyzed the available *flow* capacity ratio for the two flows that historically contribute the most traffic in each sector. Thus, the actual and the predicted (LAT=30, 60, 90, and 120 minutes) weather impact indexes (2-D weather coverage, 3-D WAAF coverage, and available flow capacity ratio) were collected for each 15 minutes of June and July 2007. WAAF-A8 was used to calculate the predicted and actual 3-D WAAF coverage and the available flow capacity ratio. The coefficients of determination (R^2) between the actual value and the predicted value for each LAT were calculated for each sector weather impact index. Here are the basic conclusions from this initial analysis:

- Predictability of 2-D weather coverage and 3-D WAAF coverage are comparable.
- 3-D WAAF coverage and the reduced flow capacity ratio tend to over predict due to the over-prediction of echo top height.
- Predictability of the available flow capacity ratio is relatively lower than the other two sector weather impact indexes due to its sensitivity to the accuracy of predicting the weather shape and location.

The rest of the section illustrates these conclusions with examples. The first example is ZID16, with two major flows, ZDC36-ZDC16-ZDC10 and ZDC36-ZDC16-ZDC12 shown in Figure 3-15. Figure 3-16 shows the predictability results for ZDC16. The R^2 between the actual value and the predicted value for each weather impact index is plotted as a function of LAT. For the available flow capacity ratio, the R^2 for each major flow is plotted. Note that R^2 decreases as LAT increases for all weather impact indexes. The predictability of 2-D weather coverage (the blue line) and 3-D WAAF coverage (the green line) are comparable. But the predictability of the available flow capacity ratio (the red and purple lines) is lower than the other two.

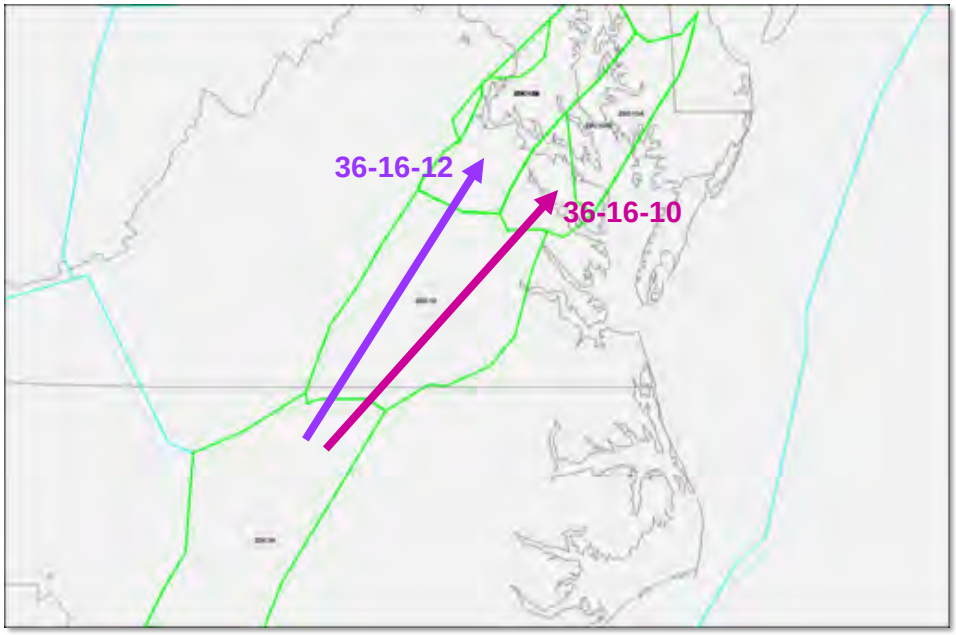

Figure 3-15. Top Two Major Flows of ZDC16

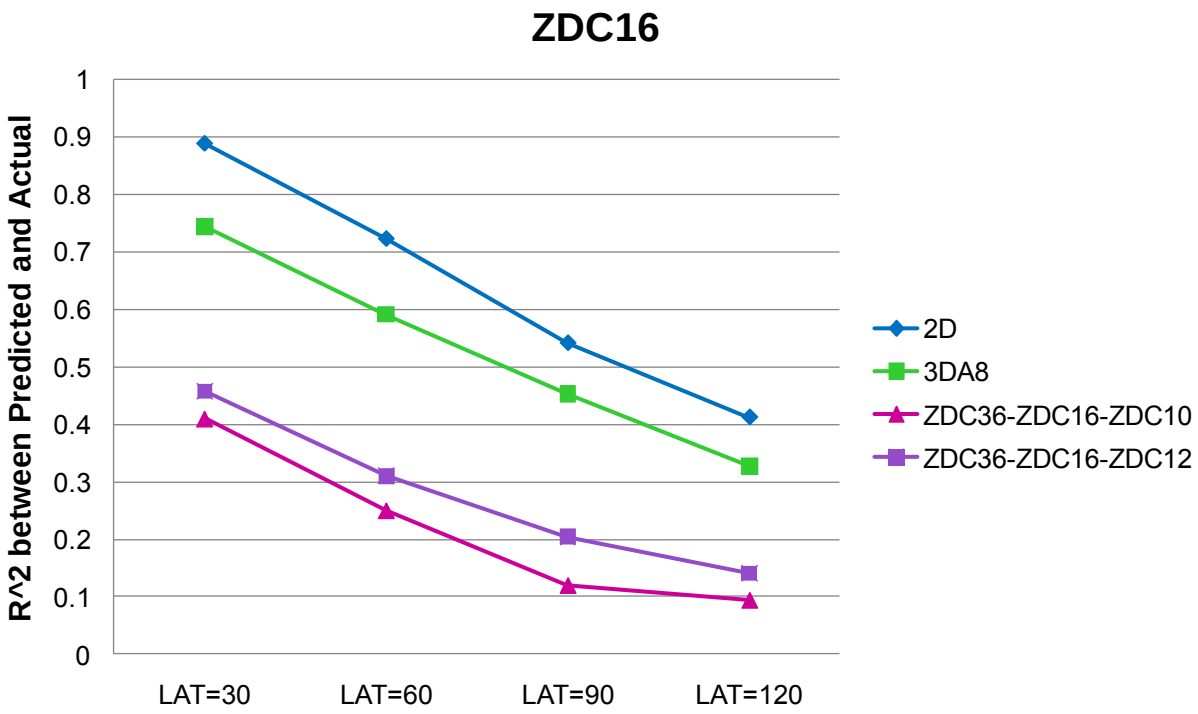

Figure 3-16. Predictability of Weather Impact Indexes for ZDC16

The low predictability of the available flow capacity ratio is due to the sensitivity of the available flow capacity ratio to the location and shape of the weather area. Figure 3-17 shows two cases with the same size and shape of weather in different locations within Sector ZDC16. The available flow capacity ratios for flow ZDC36-ZDC16-ZDC12 in case A and case B are totally different. For case A, the mincut without the weather (Omincut) is equal to the mincut with the weather (Wmincut), so the available flow capacity ratio is 1. But for case B, Wmincut is zero since there is no way to get through the weather area within ZDC16, so the available flow capacity ratio is zero.

Figure 3-18 shows two cases with weather of the same size but different shapes at the same location within ZDC16. Again, for case A, the available flow capacity ratio would be 1 because Omincut = Wmincut; and for case B, it would be zero because Wmincut is zero.

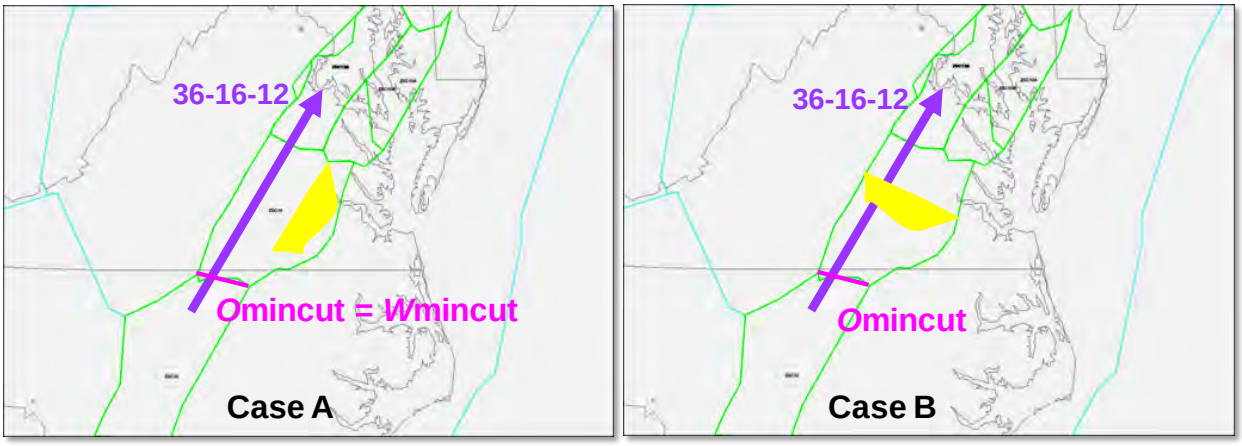

Figure 3-17. Sensitivity to Weather Location

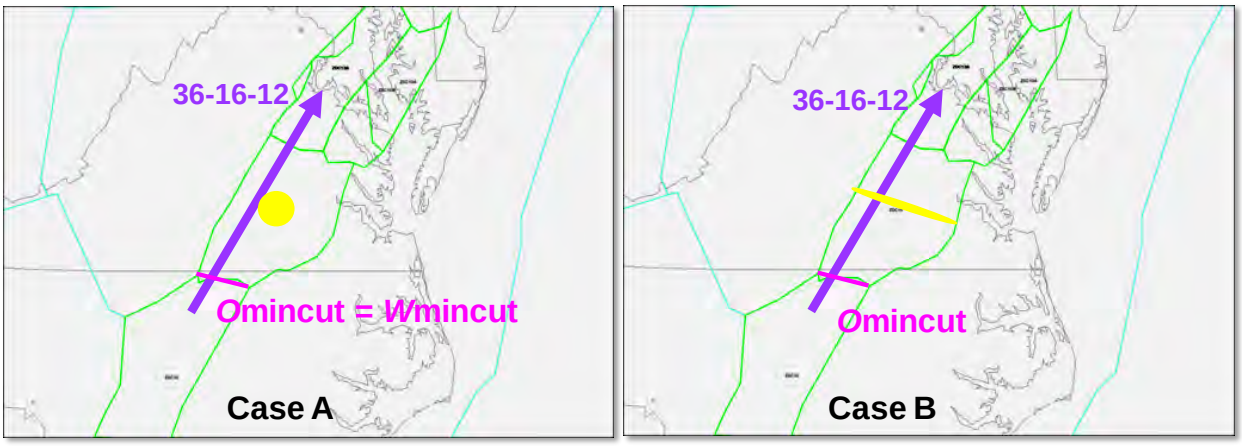

Figure 3-18. Sensitivity to Weather Shape

The second basic conclusion from the linear correlation analysis between the predicted and the actual weather impact indexes is that both 3-D WAAF coverage and the reduced flow capacity ratio (1 - the available flow capacity ratio) tend to over-predict. Figure 3-19 shows a scatter plot of predicted and actual observations of 2-D weather coverage and 3-D WAAF coverage calculated with WAAF-A8 for LAT=30 minutes in ZDC16. The green line is a linear fit to the data, and the black line is the **x=y** line where **x** is the actual observation of the weather impact index and **y** is the prediction at 30 minutes before that actual observation. As we can see from the figure, most of the points on the 3DA8 plot are above the **x=y** line, which means the predicted value is larger than the actual observation. Since both weather impact indexes use CWAM1, which considers echo tops in addition to VIL, the over-prediction of the 3-D WAAF coverage and the reduced flow capacity ratio should be due to the over-prediction of echo tops.

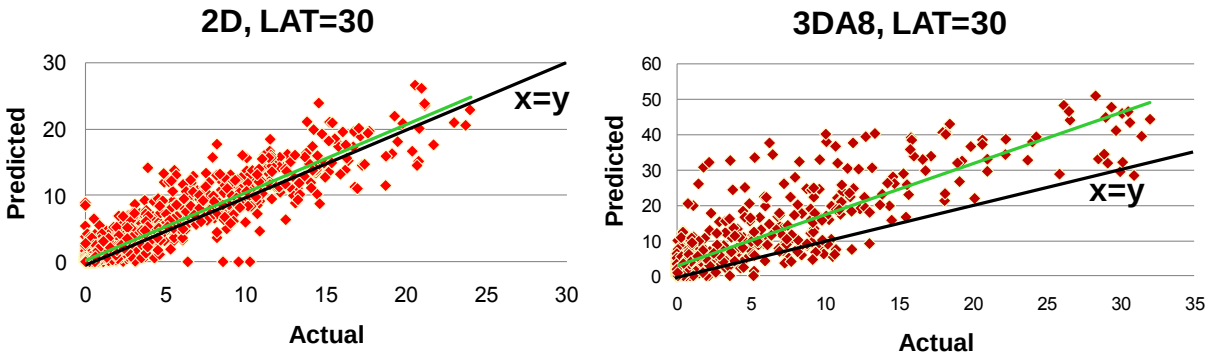

Figure 3-19. Evaluating Weather Metric Bias for ZDC16

The predictability of the sector weather impact indexes also depends on the location of the sector. The predictability of the sector weather impact indexes for the sectors in ZOB, for example, is very low due to the low predictability of weather near the Great Lakes. Figure 3-20 shows the predictability of the sector weather impact indexes for ZOB77. As we can see from the figure, the available flow capacity ratio is not predictable at all for ZOB77. The predictability of 2-D weather coverage and 3-D WAAF coverage for ZOB77 are also much lower than sectors in other centers.

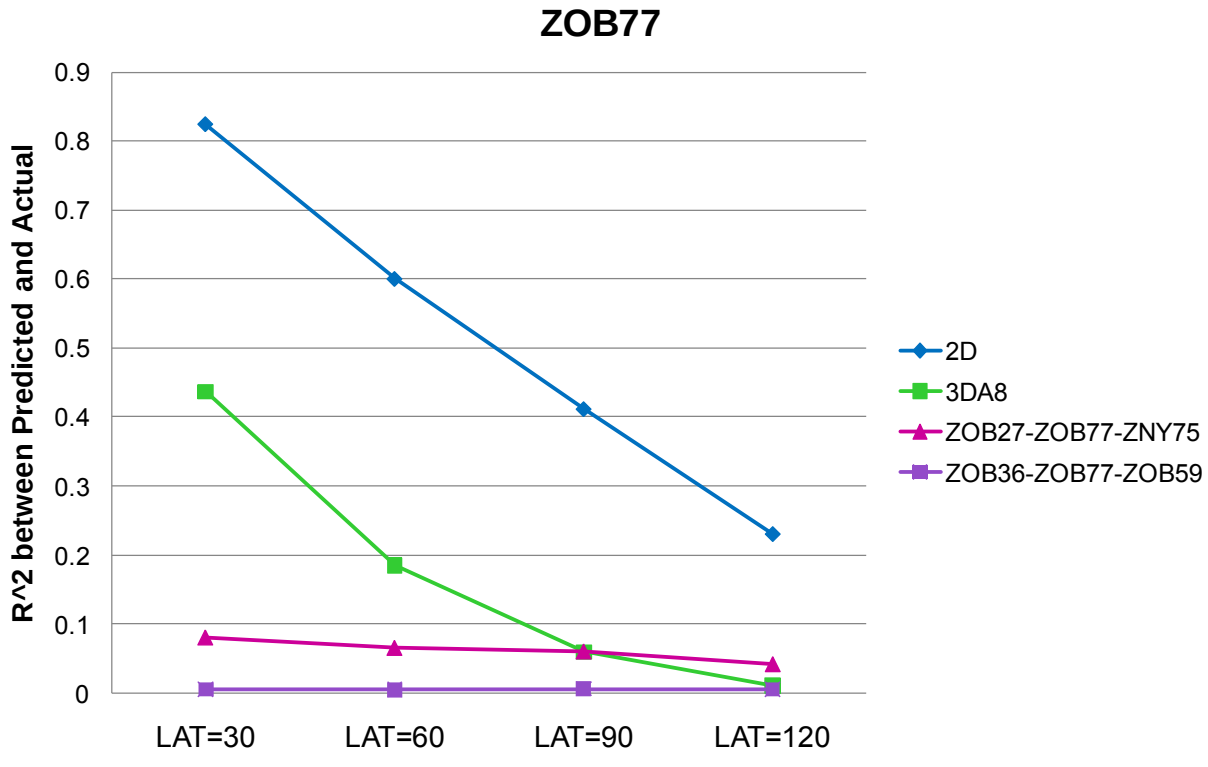

Figure 3-20. Predictability of Weather Impact Indexes for ZOB77

Of the sectors we examined in four centers (ZDC, ZNY, ZOB, and ZID), the weather impact indexes for ZID sectors have the highest predictability. Figure 3-21 shows the predictability of weather impact indexes for ZID66. As shown in the figure, the available flow capacity ratios of the top two major flows in ZID66 can be predicted up to 90 minutes if predictable is defined as $R^2 > 0.3$.

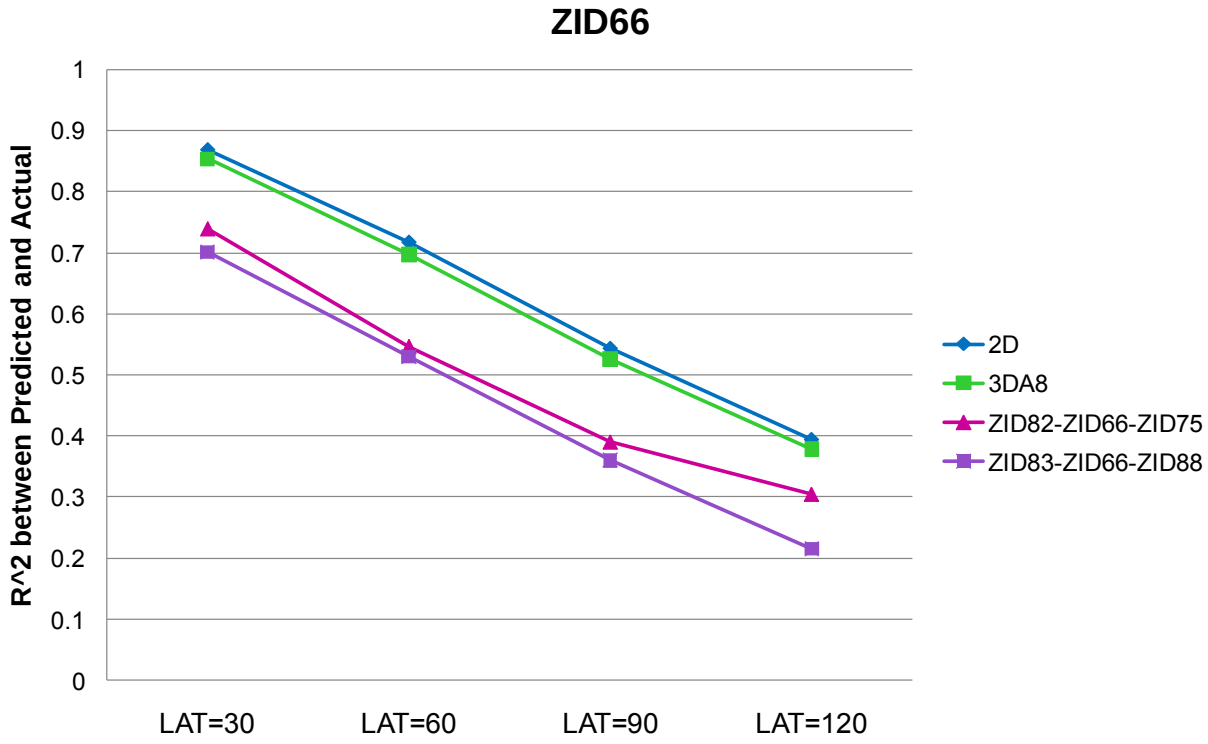

Figure 3-21. Predictability of Weather Impact Indexes for ZID66

3.5 Summary and Next Steps

In this task, we evaluated three models for predicting sector capacity under weather impact. The three models introduced three sector weather impact indexes: 2-D weather coverage, 3-D WAAF coverage, and the flow-based available sector capacity ratio. The linear correlations between the three sector weather impact indexes and the estimated actual sector capacity were examined for 48 high sectors and 43 low sectors from four centers (ZDC, ZNY, ZOB, and ZID). The flow-based available sector capacity ratio has the strongest linear correlation with the estimated actual sector capacity for most of the high sectors with some types of WAAF. For low sectors, 2-D weather coverage did as well or better than the other weather impact indexes.

Although the flow-based available *sector* capacity ratio has the best linear correlation with the estimated actual sector capacity, the predictability of the available *flow* capacity ratio is relatively low compared to the 2-D weather coverage and 3-D WAAF coverage, due to its sensitivity to the location and shape of the weather area, which are much harder to predict than the weather size.

This limits the effective LAT at which flow-based methods can be used, until the predictability can be improved.

There are two areas where we propose to focus the next steps of this research. First, we will establish a feasible candidate method of predicting weather-impacted sector capacity for TFMS WP2. The research presented here provides a basis for an initial model, in which: (1) low sector capacities are predicted using a 2-D coverage model, (2) high/super-high sector capacities are predicted to LAT <= 45 minutes using a flow-based model, and (3) high/super-high sector capacities are estimated with a 2-D model at LAT > 45 minutes. Recall that we have determined correlations between weather impact and sector throughput, but have not established capacity values. To apply the models operationally, they need to be calibrated to observed sector throughput under high-demand situations to establish weather-impacted capacities. Methods for updating the calibration over time also need to be established. In addition, when new sectors are introduced, there will be no historical data with which to calibrate them, so a method for setting initial capacity parameters needs to be developed.

The second area of research is focused on a longer-term solution. The studied models do not consider two factors that could be important, namely how to estimate weather impact on climbing or descending flows, and how to capture the impact of traffic complexity on sector capacity when weather is present. We plan to study these factors and propose ways of addressing them in the flow-based model. Also, weather products continue to change and improve, and may provide opportunities to improve capacity prediction. For example, the National Center for Atmospheric Research (NCAR) is developing probabilistic ensemble forecasts that may support probabilistic sector capacity prediction [Steiner et al., 2009]. Finally, we plan to continue collaborating with MIT/LL on their evolving CWAM model to obtain better predictions of weather blockage, such as in low altitude sectors.

4 Flight Options Generation for Semi-Automated Congestion Resolution

Flight-specific congestion resolution methods share one common requirement; the need to evaluate alternative trajectory options for flights involved in predicted congestion situations. These options include ground delays, reroutes, and altitude constraints, and may include required-time-of-arrival constraints at specific flight waypoints in the NextGen timeframe. Most existing flight-specific congestion resolution algorithms, including those discussed in Sections 5 and 6, use two primary sources for options. First, customer-provided options are considered, if available. If not, a database of pre-coordinated or historically-used routes is searched to find acceptable alternatives, with or without ground delay (altitude changes are typically not yet considered). In FY09, research was done both to improve the option database and to develop methods of generating acceptable new routes in real-time. Both the new databases and the route generation algorithm have been implemented in the En route Flow Planning Tool (EFPT) prototype [Bateman et al., 2009] for evaluation.

4.1 Route Database Enhancements

Our route alternative databases are generated from Coded Departure Routes, Playbook Plays, and historically flown routes. We generated the databases from 350 days of historical data containing 2.9 million routes as filed at the time of departure. Three different databases have been generated: Origin to Destination, Fix to Destination, and Fix to Fix.

The Origin to Destination database consists of complete routes from departure airports to arrival airports. For example, a query for all routes from Boston's Logan International Airport to San Diego International Airport yields 178 different routes. Each route record contains the identifiers of the departure and arrival airports, the route string, the distance in nautical miles of the route, and historical usage counts for the route. This database is used to find candidate reroute options for an inactive flight by the origin and destination of that flight. Programs that use the database can reject routes with very low usage counts since they may not be operationally acceptable in most situations.

The Fix to Destination database consists of partial routes starting at named fixes and proceeding to destination airports. This database is created using the Origin to Destination database. For example, a query for all routes from Fix BRIBE (near Chicago) to Atlanta Hartsfield International Airport (ATL) yields 30 different routes. Each route record contains the starting fix name, the destination airport, the route distance from the starting fix to the destination, the usage count, and the route string. This database is useful for finding route candidates for active flights, by looking for a fix along the current filed route where the flight can turn off its route and then use a new Fix to Destination route to find options for continuing to its destination.

The Fix to Fix database consists of route segments from one fix to another, where the two fixes are consecutive fixes in one or more routes from the Origin to Destination database. Each record of the database contains the two fix names, the distance between them, the heading from the first fix to the second, and the combined historical usage of this route segment of all the historical

routes that used the segment. For example, there is a record for BRIBE to Fix COTON where the segment between these fixes was used 16979 times. There is also a record for COTON to BRIBE where the segment was only used 1097 times, indicating this segment is usually used in one direction. All of the Fix to Fix database records combined form a route segment network over the whole NAS. The Fix to Fix database cannot yield reroute candidates directly, but can be used by route generation algorithms to help generate new ad hoc routes.

The Origin to Destination and the Fix to Destination databases are being used directly in the EFPT prototype to find candidate reroutes for inactive and active flights respectively. The EFPT prototype also uses the Fix to Fix database for its Initial Ad Hoc Route Generation process, which is described in Section 4.2.

The FY09 improvements to these databases included speeding up query access to the databases. FY09 software changes and database management manipulations of these databases have improved query speeds by at least a factor of 3. But query speed is still an issue, because applications like EFPT can produce thousands of queries when generating a solution, and most of the time spent generating solutions is spent in querying the databases.

4.2 Initial Ad Hoc Route Generation

An initial ad hoc route generation capability has been created based on database queries into the Fix to Fix database. The intention of this capability is to provide an early capability that could be used in EFPT and other applications. This capability could be replaced when other ad hoc reroute generation research becomes mature enough and fast enough to use.

This capability queries the Fix to Fix database using the name of the first fix, a range of headings, and a range of distances. Figure 4-1 shows an example where a flight has filed a route that includes the fixes ARUBA and LUCIA (fix names in this example are fictional) and there is weather predicted to block that part of the route. The ad hoc route generating algorithm will begin building routes by querying the Fix to Fix database to find nearby fixes that are generally in the right direction and may be able to avoid the weather. The query selects segments that start at ARUBA, are within a min and max distance range, and have headings within a range that is centered on the line from ARUBA to LUCIA. The query area looks like the area swept by a windshield wiper. In this example, the query produces three results: KITTS, MARIN, and NEVIS.

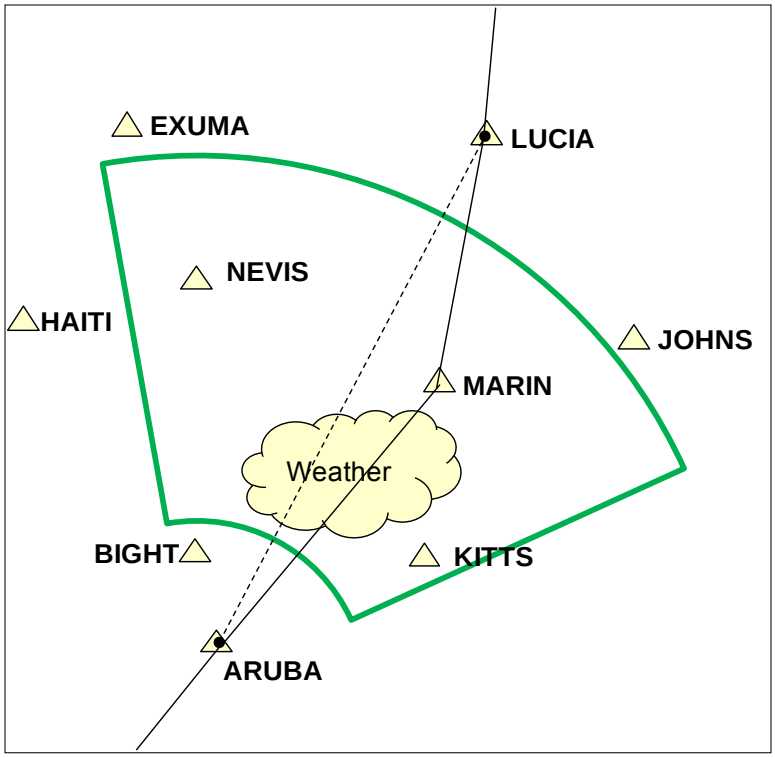

Figure 4-1. First Query to Find Fixes

The route generation algorithm then steps to KITTS and a new query is launched to find fixes that are in the direction of LUCIA from KITTS, as shown in Figure 4-2. This query produces MARIN, JOHNS, and LUCIA.

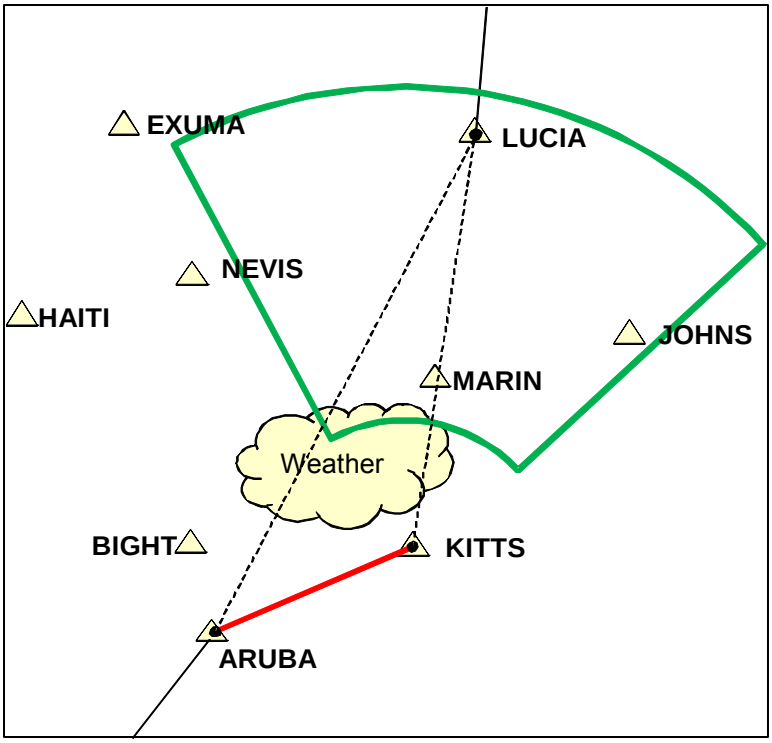

Figure 4-2. Second Query to Find Fixes

The segments to MARIN and to LUCIA are both blocked by weather, so they do not need to be examined further. JOHNS is not blocked, so the algorithm steps to JOHNS and issues a new query shown in Figure 4-3. The distance range is smaller here, because there is no need to query distances that are much longer than the distance from JOHNS to LUCIA. LUCIA is the only result from this query. Since the goal fix has been reached, the first candidate reroute is complete: ARUBA.KITTS.JOHNS.LUCIA. A new route string can be formed for a flight by substituting this route into their route string between ARUBA and LUCIA.

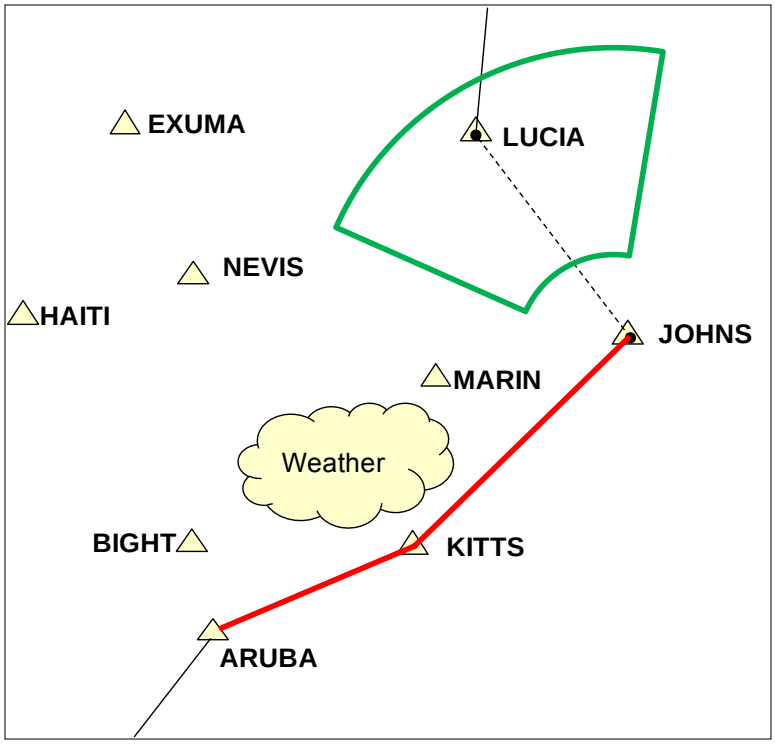

Figure 4-3. Third Query to Find Fixes

The algorithm continues the search for candidate routes by examining other fixes from earlier queries. This is basically a depth-first tree search. The first query, shown in Figure 4-1, also produced NEVIS, and a query from NEVIS will produce LUCIA and form another candidate reroute: ARUBA.NEVIS.LUCIA.

Care must be taken to limit the size of the search space, since it can grow geometrically. Features are included in the algorithm to help limit and prune the search space, including:

- Segments are checked for weather impact before they are followed.
- There is a turn angle limit for each fix. This is factored into the heading range when queries are issued.
- There is a maximum and minimum distance allowed for each route segment.
- There is a maximum number of fixes allowed in the new candidate reroute segment (from ARUBA to LUCIA in this example).

This algorithm is used by the EFPT prototype to help find candidate reroutes. In some cases all of the Origin to Destination and Fix to Destination routes are impacted by weather and ad hoc routes are the only candidate reroutes. Also, ad hoc routes can lead to better grouping where larger numbers of flights are able to use a candidate reroute. The reason is that these routes can often handle flights with many different destinations. With the Origin to Destination or the Fix to Destination databases, all flights on any one candidate reroute must be going to the same destination.

This capability has been fast enough to use in the initial demonstrations, since we can limit the demonstration story to a few route blockages in order to control the time needed to generate reroute candidates. But processing speed has been an issue because larger problems can take over a minute to solve and there is a desire to expand the search by trying earlier and later fixes in the route in order to find more candidate reroutes. An examination of the processing showed that the majority of the time was being used by the database to process queries.

Since the Fix to Fix database is small enough to fit into memory (about 500 megabytes) we wrote a special memory resident database optimized to handle only the limited types of queries made by this route generation algorithm. This cut the overall processing time for EFPT problem solving in half. This means the database access times are much more than twice as fast as before, since the overall processing time includes other processing such as queries to the Origin to Destination and the Fix to Destination databases.

4.3 Next Steps

The flight option generation process described here will be tested as part of the EFPT evaluation process [Bateman et al., 2009]. Feedback from subject matter experts will be used to refine the search space and parameters so that more useful options are generated.

As mentioned in Section 1.3, there is a parallel research project, *Flight Option Generation (FOG) for NextGen Automation*, that will continue in FY10. It is expected that the FOG project will provide a more general purpose capability for generating options with desired operational characteristics. This capability will also produce quantitative metrics for the options it provides (such as consistency with established flow patterns, potential weather impact, delay) to be used in choosing the best option for a specific congestion management problem.

5 Continual, Probabilistic Congestion Management

In FY08, we developed a sequential, probabilistic congestion management approach that accounted for both traffic demand and convective weather forecasting uncertainty [Wanke et al., 2008a]. That approach formulated the management of predicted congestion at a fixed future time as a decision-tree problem, where sequential congestion resolution actions of varying aggressiveness are used to resolve congestion before the predicted congestion time is reached.

Simply defined, en route congestion exists when the predicted traffic demand exceeds the predicted capacity of one or more NAS air traffic control sectors. Today, automation provides a deterministic prediction of traffic demand, and this is compared to fixed sector capacities to provide potential congestion alerts. The effect of weather, and of uncertainty in weather and traffic demand forecasts, must be estimated by traffic managers. Given this, traffic managers typically try to match the demand to the capacity. If the predictions (automated or mentally-projected) of demand and capacity are unbiased, then this implies that the traffic management objective is to achieve a 50% probability of congestion.

So if we have a quantitative probabilistic congestion forecast with a congestion probability greater than 50%—perhaps using the techniques described in Sections 2 and 3—then a decision about congestion management needs to be made (Figure 5-1). How much control is needed to ensure that the predicted congestion is resolved before the problem time is reached? This decision is made knowing that the strategy will be modified at the next decision time, and thus the problem need not be completely solved now. If too-aggressive action is taken, then some flights will be affected unnecessarily. Also, weather and traffic predictions for the time of interest will improve if we wait to act. On the other hand, if insufficient action is taken early, then more intrusive maneuvers, such as longer reroutes or reroutes to airborne flights, may be required to manage congestion.

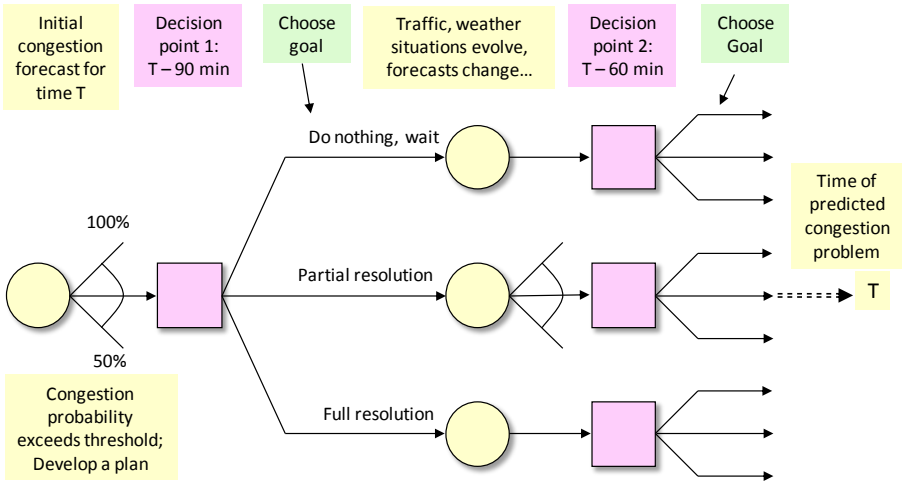

Figure 5-1. An Abstract Sequential Decision Tree for Congestion Management

In the previous work, the abstract strategies described in Figure 5-1, such as "partial resolution," were transformed into quantitative goals for reducing the predicted probability of congestion to a designated target (that is, of demand exceeding capacity during the predicted congestion event). We carried out a Monte Carlo simulation study, using models of both traffic and weather forecast uncertainty, to determine the "best" strategy for three different levels of weather forecast uncertainty. This was done by traversing all possible paths through the decision tree and comparing the results in terms of congestion resolution effectiveness and delay. For the scenarios tested, the best strategy involved a partial resolution at the beginning of the resolution process, transitioning to "full" resolutions as the predicted congestion time approached. When weather forecast uncertainty increased, the cost (in terms of flight delays) of solving the congestion also increased, as well as the variability in cost across the Monte Carlo simulation outcomes.

In FY09 we have converted the best sequential strategies for a fixed prediction time to continual resolution strategies [Wanke, 2009], in which congestion maneuvers are computed periodically, using a congestion risk goal that varies with prediction LAT. This represents an automated version of the "Develop Risk Management Strategy" box in Figure 1-1 (reproduced here as Figure 5-2). A modified version of the simulation was constructed to test several candidate strategies. Since these strategies do not require computation of an entire decision tree, they are adaptable to real-time decision support, provided weather forecast uncertainties can be suitably quantified.

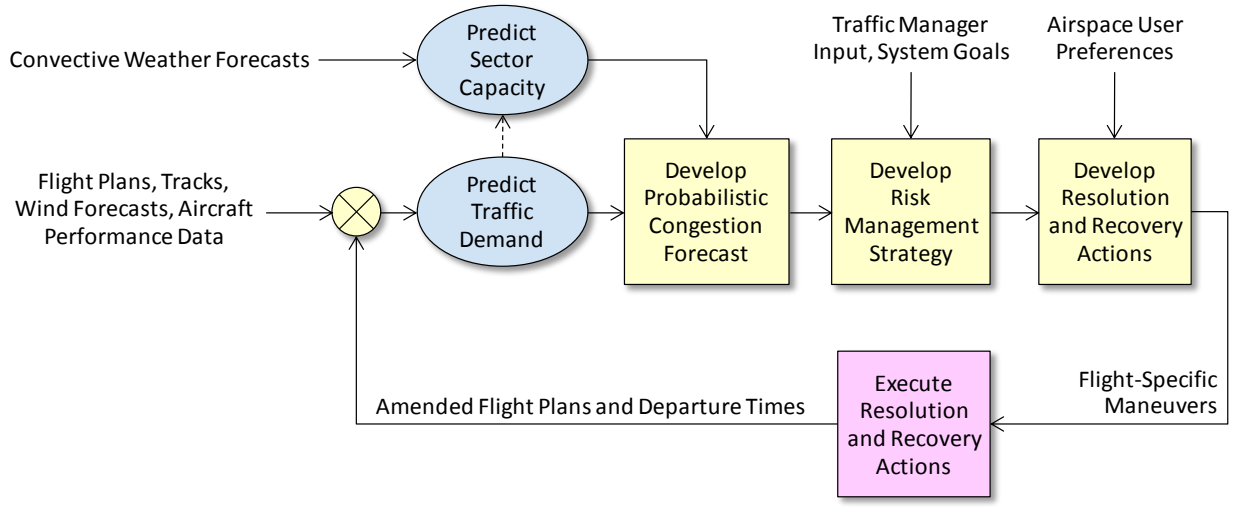

Figure 5-2. A Continual, Probabilistic Congestion Management Concept

5.1 Quantitative Risk Management Goals

To implement a risk management strategy, the congestion forecast and resolution goals must be expressed in quantitative terms. We have chosen to use "congestion probabilities," meaning the probability that the demand will exceed the predicted capacity in a given sector at a given time. Resolution strategies can then be specified in terms of target congestion probabilities or "congestion risk," which may vary as a function of prediction LAT. If the traffic demand and

sector capacity predictions can be generated such that they are PMFs[15] rather than point estimates, then these predictions can be combined through convolution to determine the probability that the demand will exceed the sector capacity at each prediction time. This is shown in Figure 5-3 for a single sector, via box plots of predicted distributions of demand and capacity for a single sector over a series of 15-minute intervals.

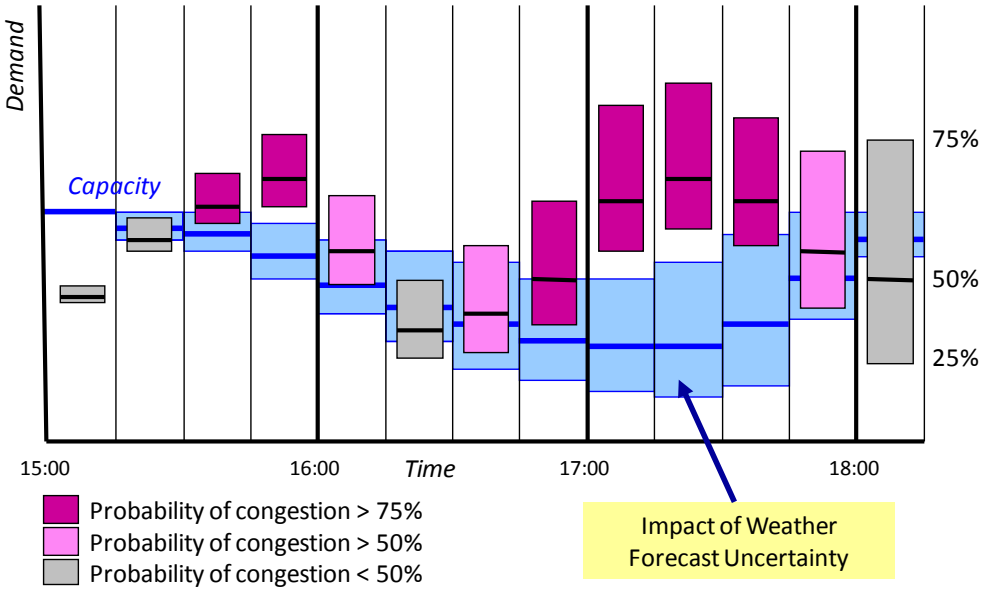

Figure 5-3. Predicting Probability of Congestion Using Uncertainty Distributions for Traffic Demand and Sector Capacity

The heavy blue lines represent the 50th percentile prediction of airspace capacity, surrounded by a blue box representing the 25th to 75th percentile prediction range (a standard box plot). The capacity is well-known initially, since weather is not predicted to impact the area until, at earliest, 15 minutes into the future (1515). The clear weather capacity has traditionally been assumed to be the MAP [Volpe, 2002] used by current NAS traffic management automation.[16] At greater LAT, the weather is expected to reduce capacity, and the spread in possible values of that capacity reflects the uncertainty in the future position, size, and intensity of the weather. Similarly, the demand data (gray, pink, and purple boxes) increase in uncertainty as LAT increases. The demand boxes are color-coded by the probability that the demand will exceed the

---

[15] In this work, capacity and demand are expressed in terms of aircraft count, an integer quantity, thus the distribution is expressed as a PMF rather than a PDF.

[16] MAP is expressed in terms of maximum instantaneous sector count within a single sector, and values are assigned to all en route NAS sectors via a formula with adjustment by operational personnel. Values range from 6 to 25, depending primarily on sector size and secondarily on traffic complexity. MAP is intended to be an alerting criterion, not a capacity, since traffic managers can allow MAP to be exceeded if traffic conditions indicate; but since demand that exceeds the MAP triggers traffic management personnel to analyze the problem, it makes a reasonable stand-in for capacity at present.

capacity (that is, there is congestion) during each interval. There are two periods (1530 - 1559, 1645 - 1744) during which the probability of congestion is greater than 75%.

Once such a prediction is available, flight-specific maneuvers such as reroutes or ground delays can be selected to modify the traffic demand and reduce congestion risk. Figure 5-4 shows the effect of reducing the congestion risk to a particular goal profile; no more than 50% for LAT up to 60 minutes, and no more than 75% for LAT between 60 and 120 minutes. This reflects the results of the prior study, which suggested that partial resolutions were appropriate for longer LAT. Note that a 50% target is considered ―full resolution," since it roughly matches the demand to the capacity; however, a conservative approach could conceivably use an even lower target. For the work presented here, a relatively simple method is used to select flight maneuvers to achieve the risk profile in a computationally-efficient way [Wanke and Greenbaum, 2008], though more optimal methods are certainly feasible.

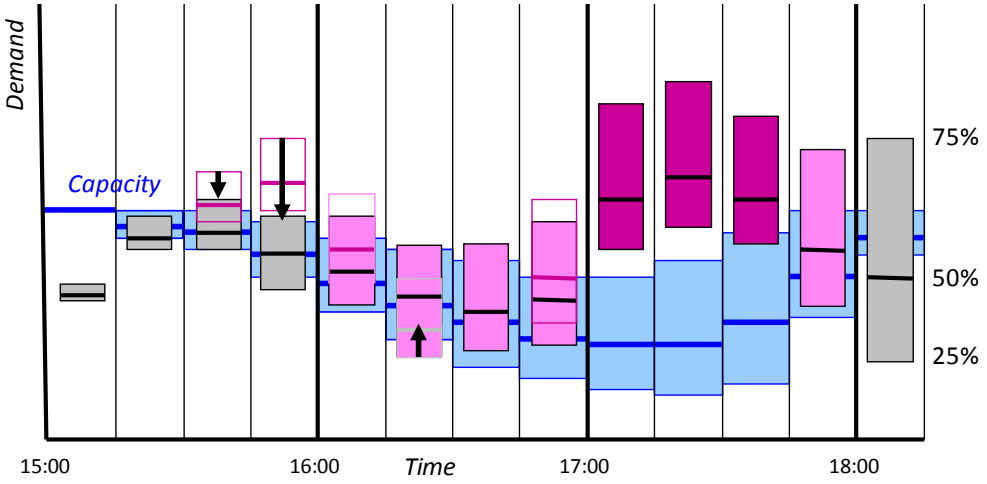

Figure 5-4. Reducing the Probability of Congestion to a Target Risk Profile

The challenge, however, is to select the appropriate target congestion risk profile to achieve the proper balance between incurring unnecessary delays and failing to resolve the congestion. Thus, we postulate a range of strategies from ―wait-and-see" to ―aggressive," simulate them, and evaluate the results in terms of congestion resolution effectiveness and incurred delay. Defining the ―best" balance between these metrics is difficult, since it is a mathematical representation of a traffic management objective, which has traditionally been defined qualitatively.

Even if an effective congestion risk profile is chosen, there will be some cases where the weather turns out to be significantly less severe than forecasted. In these cases, we would like to "undo" previous maneuvers (delay recovery) to the extent possible. The degree to which maneuvers can be undone depends on the current state of the flight. If a reroute was issued to an airborne flight, or to an inactive (pre-departure) flight that has since departed, it may be problematic or even undesirable to return them to their original clearance. If the flight has not yet departed, then removal of a reroute is generally feasible, and some or all of an assigned ground delay may also be removable, depending on the status of passenger loading, luggage, etc. In this concept, we

allow for delay recovery when doing so will not raise the congestion risk above a recovery congestion risk profile. This profile could be the same as that used for resolution maneuver generation, or have lower risk values, but can never have higher values; otherwise, flights would be repeatedly maneuvered and recovered, even if the weather and traffic predictions are correct.

5.2 A Monte Carlo Simulation Method for Evaluating Continual Congestion Management

Tactical congestion management is a complex process, and complex to simulate. We are attempting to control aggregate quantities, specifically, balancing traffic loads against available capacities while minimizing delay and schedule disruption. However, we would like to control traffic loads efficiently, which requires flight-specific maneuvers (ground delay, rerouting). Therefore, we must simulate predictions and actual outcomes for individual flights. This requirement rules out traditional closed-form methods such as dynamic programming, which require modeling the system as a Markov process. Thus, a Monte Carlo simulation approach is used.

The essential tradeoff in sequential decision making under uncertainty involves flexibility versus knowledge. If we wait to make a decision, we will learn more about the possible outcomes and thus be able to make better decisions. However, some actions that were available earlier will no longer be available. For example, it is easier and less costly to ground delay a flight than to reroute it after it departs. So, a simulation of this concept must capture both the range of possible outcomes and the increase in knowledge as time progresses.

5.2.1 Overview

Our method begins by defining the congestion problem as (1) a set of trajectory predictions for traffic demand, (2) a nominal set of clear-weather sector capacities, and (3) an initial convective weather forecast. Next, we define the risk management strategy: a set of target congestion risks as a function of LAT, and how often the situation will be reevaluated for new maneuvers. The resolution strategy also includes constraints on what kind of maneuvers are permitted, for example, whether or not we will allow rerouting of airborne flights to solve congestion. Finally, we define the conditions under which delay recovery is possible, and the target congestion risk for delay recovery actions.

The simulation is started by running a Monte Carlo simulation for both traffic and weather outcomes. The weather outcomes are converted to sector capacities using an empirical model based on the weather impact metric described in Section 3.1.1. This provides the key advantage of this approach over the study of historical severe weather events, which can only have a single outcome. With the Monte Carlo simulation, we can explore the whole reasonable range of outcomes for a single scenario, and therefore determine what the best congestion resolution over the entire range of outcomes would be. Since historical weather events are as different as snowflakes, it is nearly impossible to gain statistical data on the goodness of a particular strategy by studying historical events.

At each time a decision is to be made, referred to as a decision point, the resolution strategy is applied. Resolution maneuvers are computed to achieve the desired congestion risk profile (as in Figure 5-4) via a heuristic assignment method, based on closed-form statistical models of the uncertainty in demand and capacity predictions. These models are different from the Monte Carlo outcome models, since they must be computable in real-time. Rather than enumerating a series of outcomes, they compute the statistics of the aggregate prediction error in aircraft demand count and in sector weather coverage fraction. The weather coverage fraction is then converted to capacity using an empirical model. The resolution maneuvers are computed to achieve the target level of risk for all sectors in the desired region simultaneously, since each flight can affect several sectors within the congestion region.

Once resolution is complete, delay recovery possibilities are evaluated. Flights assigned resolution maneuvers that can still be "undone" are examined, and if the maneuver can be partially or wholly undone without exceeding the recovery congestion risk or risk profile, then the maneuver is undone. The trajectory predictions for each outcome are updated to reflect the resolution and recovery maneuvers chosen for that outcome, and the "actual" trajectory outcomes are modified to reflect the maneuvers, with some uncertainty about how they will be executed.

Time is then advanced to the next decision point. At that point, more is known in each outcome about how the actual traffic and actual weather are occurring, so the traffic demand and weather predictions are updated to reflect the knowledge gained.

This process continues until the last decision point in the scenario is reached. Metrics are computed at each decision point and cumulatively to describe the impact and success of the resolution strategy. The following subsections describe the details of the simulation process.

5.2.2 Modeling Traffic Outcomes

The initial traffic prediction comprises a set of predicted flight trajectories and departure times. Because we wanted to develop flight-specific resolution actions at each decision point, we developed a Monte Carlo model to simulate the possible "actual" flight trajectories that would be flown given such a traffic prediction [Hoffman et al., 2007]. It models the following, for predicted flights:

- Cancellations
- Departure time estimation errors
- Changes in route and cruise altitude
- Flight progress estimation (speed) errors

Also, the model will create and add a set of flights that have not filed at the time of the prediction but will appear before the time for which the prediction was made ("pop-ups," also discussed in Section 2.1). Thus, for a single traffic prediction, the model will create N different outcomes. Each of these outcomes may contain a different number of flights (due to cancellations and pop-ups), and the flights will differ from the prediction due to the other prediction error distributions. These distributions were developed empirically, based on current prediction methods and

procedures. If future ATM concepts are to be modeled, and these concepts would change the predictability of trajectories, the model distributions could be adapted to reflect that.

5.2.3 Sector Capacity Modeling

Before developing the Monte Carlo model for weather outcomes, we needed to understand how weather parameters can be related to what we actually need for the congestion calculation, namely, future sector capacity outcomes. This study used the two-dimensional weather coverage metric described in Section 3.1.1, namely, that the percentage of a sector covered by VIL3+ precipitation is linearly related to the effective sector capacity. We calibrated this metric based on the empirical observation that sector capacity becomes effectively zero as the coverage fraction increases beyond 50%. Also, sectors are assumed to have a peak aircraft capacity equal to the MAP when the weather is clear. This capacity value is assumed to fall linearly with increasing coverage fraction until the coverage fraction reaches 50%, and at 50% and above, the capacity is assumed to be zero.

This is a large simplification, since the position, shape, and height of a weather event are important in determining where capacity will be lost. Other research has proposed airspace capacity models that consider flow patterns ([Song et al., 2007b]; also see Section 3) and route organization [Martin, 2007]. We may use such models in follow-on work.

5.2.4 Modeling Weather Outcomes

Given this capacity model, we chose to represent weather as sector coverage fraction. Weather predictions are assumed to take the form of predicted sector coverage fractions as a function of look-ahead time. These time series are created for each sector of interest in the simulation, and referred to as ―coverage traces.‖ A sector coverage trace can be converted to a sector capacity trace, via the algorithm described in Section 5.2.3. So, the Monte Carlo model needs to start with a nominal coverage trace and generate a set of ―actual‖ coverage traces which represent a range of weather outcomes.

As noted earlier, some current weather forecast products contain measures of uncertainty. However, these are not easily adapted to computing the uncertainty in predicting sector coverage fractions. Thus, a simplified method of quantifying coverage uncertainty was adopted. Prediction uncertainties in the tactical TFM timeframe are assumed to occur in two independent forms: errors in predicting weather evolution speed, and errors in predicting the growth and decay of weather intensity.

Nominal traces for each sector are created either from an observed weather event or synthetically, to represent a desired capacity reduction scenario. These traces are taken as the initial weather prediction for the scenario. Then, a Monte Carlo set of ―actual‖ weather outcomes are generated by first computing speed and intensity variations from closed-form statistical distributions, and then dilating and scaling the nominal coverage traces to represent the ―actual‖

coverage traces.[17] This process is illustrated in Figure 5-5. The blue trace represents the nominal weather coverage, the red indicates weather which is both faster to evolve and more intense than predicted, and the green trace indicates weather that is less intense and slightly slower to evolve than expected.

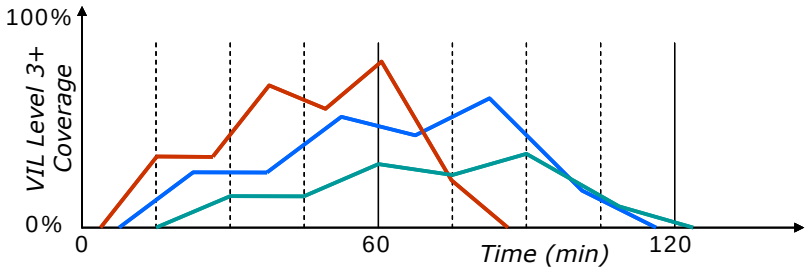

Figure 5-5. Speed and Intensity Variations on a Weather Coverage Trace

In the absence of empirical data on this process, which we plan to obtain and use in future work, we employ independent, zero-mean, symmetric triangular distributions for speed and intensity prediction error. The half-width of the distributions are varied to represent different levels of uncertainty. In reality, these and other parameters would be functions of the weather prediction model, storm type, etc., if indeed this is a reasonable way to represent prediction error in the tactical timeframe. For this study, we used two uncertainty levels: moderate uncertainty and high uncertainty. Moderate uncertainty parameters were set such that the maximum evolution speed error was ±25% and the maximum growth rate error was ±12.5%. The high uncertainty parameters were set at ±50% and ±25%, respectively.

5.2.5 Modeling Prediction Uncertainty for Congestion Resolution

As noted above, closed-form aggregate models for traffic demand and weather prediction uncertainty are needed to calculate the congestion resolution maneuvers. For traffic demand, we used an earlier variant of the ADM described in Section 2 [Zobell et al., 2005; Wanke et al., 2005a]. The model forecasts peak traffic demand distributions based on four variables: the look-ahead time, the deterministic predicted peak count, the number of airborne flights in the peak count prediction, and the primary sector traffic type (departure, en route, arrival, mixed). It treats demand in all sectors as independent, which is a simplification. However, this model is very fast to compute and can be used in either simulation or real-time applications. Now that the new version of this model is complete, we will use it in future simulation studies.

For weather, we developed a similar model. It computes a distribution of predicted weather coverage error based on the nominal prediction and look-ahead time. This was developed in a different way, however, since the Monte Carlo weather model described in Section 5.2.4 does

[17] Note that speed and intensity errors are assumed to be consistent over the airspace of interest. If weather worsens earlier than expected in Sector A due to a speed forecast error, it also worsens earlier in adjacent Sector B.

not correlate specifically with a real weather forecast of known accuracy. So, we developed a model based on the overall statistics generated by the Monte Carlo simulation, using the weather pattern from the example problem in Section 5.3.1. It is based on simple, closed-form distributions of coverage prediction error, conditioned on the nominal predicted coverage for each outcome, the LAT, and the weather forecast uncertainty level (moderate or high). Both of these models predict aggregate quantities, but to be effective, they must approximate the statistics for traffic count and weather coverage that are produced by the Monte Carlo simulations. In the weather case, this is true by design. Assuming the simulations are good, they should also approximate real world statistics, and thus we are using them in parallel research into real-time decision support tools. Note that the aggregate uncertainty models cannot match the Monte Carlo model outputs exactly, such that predicted congestion risk will not match the ~~"a~~ctual" risk as computed from the Monte Carlo outcomes. This affects congestion resolution performance.

5.2.6 Congestion Resolution Algorithm

For each option at each decision point, a resolution strategy must be developed to meet the desired maximum congestion probability. The simulation uses a heuristic algorithm that can be rapidly computed [Wanke et al., 2005b], and has been shown to provide effective, though not optimal, flight-specific solutions [Nilim et al., 2003].

The resolution process begins by defining two airspaces. The first, the Congestion Resolution Area (CRA), contains those sectors identified as being congested. Flights that penetrate the CRA during the congestion period are candidates for resolution maneuvers. The second, the Congestion Management Area (CMA), is a larger group of sectors surrounding the CRA. These sectors are monitored during the resolution development process so that resolution maneuvers do not create additional congestion in the CMA.

Figure 5-6 illustrates the process. Candidate flights are subtracted from the CMA traffic count predictions. Then, the flights are placed in priority order as follows, listed from highest priority to lowest priority, and ordered within each category by arrival time to the CRA:

1. Airborne flights that had been maneuvered at a previous decision point
2. Airborne flights not previously maneuvered
3. Pre-departure flights previously maneuvered
4. Pre-departure flights not previously maneuvered

Flights that had been maneuvered at a previous decision point are placed higher in the list, to make it very unlikely that they would be maneuvered again; multiple maneuvers for a single flight cause schedule problems for NAS customers.

Next, a series of alternate route options are generated for each flight. These are selected from a database of predefined and historically-flown routes, keyed by origin-destination pair. Pre-departure flights also have the option of taking ground delays up to a set maximum value.

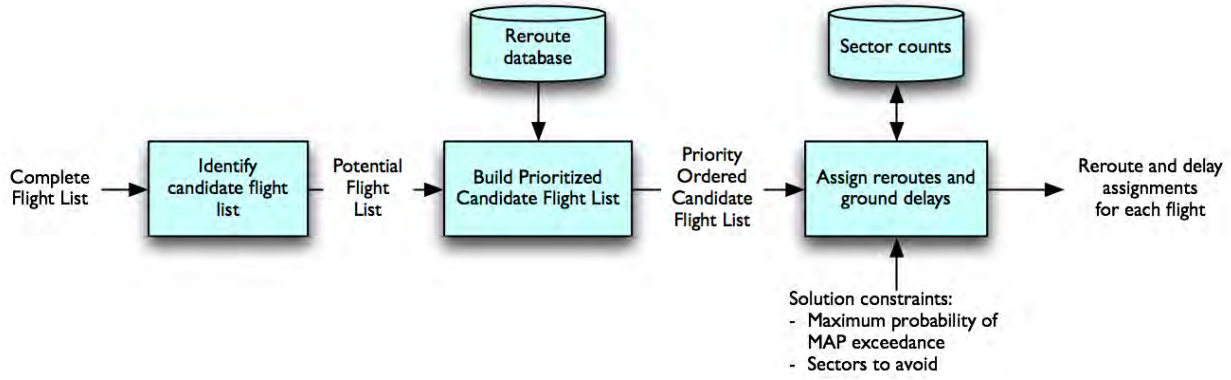

Figure 5-6. An Overview of the Heuristic Congestion Resolution Algorithm

Resolution maneuvers are assigned in a single pass through the ordered candidate list. First, the current flight trajectory is added to the CMA sector counts, and the ADM is used to evaluate the resulting congestion probabilities. If the maximum congestion risk allowed by the congestion risk profile is not exceeded, then the flight is not rerouted or delayed. If the maximum probability is exceeded, then predicted trajectories for all combinations of alternate routes and (if the flight is pre-departure) ground delays are constructed. Of the trajectories that do not violate the congestion constraint, the one with the earliest arrival time at destination is chosen as the best option. If no trajectories work, the flight is not modified, and the congestion probability goal will not be achieved. Computed maneuvers are assumed to be implemented immediately—that is, clearances for the changes to routes and departure times are issued to the flights.

Flights that are early in the prioritized list are easier to solve. As the processing reaches the end of the order, it is harder to find options that do not exceed the congestion threshold, so later flights may experience more severe reroutes and delays. This processing order is a key factor in determining the optimality and equity of a proposed solution, and it remains an area for experimentation to try other sorting approaches, or to allow additional solution options such as altitude changes.

Several parameters can be adjusted to represent different solution constraints. These include whether or not to allow rerouting of airborne flights, the number of reroute options to explore, the maximum ground delay allowable, and the minimum time before planned departure that a flight can be delayed or rerouted.

5.2.7 Delay Recovery Algorithm

The delay recovery algorithm mirrors the resolution algorithm in that it makes a list of candidate flights, prioritizes them, develops a set of recovery maneuvers for each flight, and tests each maneuver against a congestion risk profile to determine acceptability. The mechanics for each of these steps are somewhat different, however. For this study, we chose to limit delay recovery to pre-departure flights, even if the airborne rerouting option was available for resolution. This avoids the difficulty of deciding whether it makes sense to return an airborne flight to its previous route, and if so, designing a path that achieves this. Thus, we structured the algorithm as follows: A flight is a candidate for recovery if it (1) has a currently-assigned reroute or ground

delay, (2) has not yet departed and will not depart for at least A minutes, and (3) is within B minutes of its original departure time. Parameter A is a constant in this simulation, but in practice could be replaced by a value provided by the operator of that flight, such as the ~~"e~~earliest possible departure time" that NAS customers provide in the context of ground delay programs. Parameter B defines how early we are willing to start delay recovery; if B is too large, there is a risk that recovered flights might have to be delayed again.

Candidate flights are then prioritized for recovery based on the arrival time delay resulting from their currently-assigned maneuver. The flight with greatest delay is evaluated for recovery first, then on through flights with decreasing values of delay. This is only one possible ordering. One alternative would be to have NAS customers identify which flights would be most important to recover.

The best recovery option is to return the flight to its original route and departure time (full recovery). Partial recovery is also possible, by shortening the assigned ground delay (moving up the departure time), returning to the original route (if rerouted), or a combination of the two. A range of possible departure times is established by taking 5-minute intervals from the earliest to the latest possible departure time. The latest possible time is the currently-assigned departure time. The earliest is the latest of the original departure time or A minutes from the current time. A list of possible recovery actions is then made, with the full recovery option (if available) at the top, and the rest of the route/departure time combinations following in order of predicted arrival time, from earliest to latest.

This list is then evaluated in order to determine whether the recovery action would violate the recovery congestion risk profile. Since the recovery action list is sorted in order of desirability, the first action which does not violate the congestion risk profile is selected and executed. For this study, a constant recovery congestion risk goal of 0.5 was used. So, even if the resolution algorithm used tolerated a higher congestion risk at the LAT of interest, the recovery algorithm would not add flights into a sector which would eventually be resolved back to a goal of 0.5 at a later decision point.

5.2.8 Congestion Resolution Using Deferability

The deferability concept is a variation of the congestion resolution algorithm that is tailored for incremental decision making. In this concept all flights are classified as either deferrable or non-deferrable based on the time remaining until the flight is predicted to depart. For example, if the deferability parameter is set to 60 minutes, flights that depart more than 60 minutes in the future are deferrable, and all other flights are non-deferrable.

At each decision point, only non-deferrable flights are allowed to be maneuvered. Deferrable flights will become eligible for maneuvers at later decision points. This builds flexibility into the strategy by preserving resolution options for as long as possible. Additional details about the deferability concept are available in [Wanke et al., 2008a].

The congestion resolution algorithm was changed in two ways to accommodate deferred resolutions. First, the priority sort order (Section 5.2.6) is modified to include the deferrable

classification for each flight. The sort order is as follows, listed from highest priority (that is, least likely to be maneuvered) to lowest priority:

1. Non-deferrable and airborne and previously maneuvered
2. Non-deferrable and airborne and not previously maneuvered
3. Non-deferrable and pre-departure and previously maneuvered
4. Non-deferrable and pre-departure and not previously maneuvered
5. Deferrable and previously maneuvered
6. Deferrable and not previously maneuvered

Within each category, flights are ordered by arrival time to the CRA.

Second, the congestion resolution maneuvers are generated based on this priority order, but only the maneuvers for non-deferrable flights are implemented. Deferrable flights will only be maneuvered at later decision times when their time to departure becomes less than the deferability parameter time, and if it is still necessary to maneuver the flight to meet the congestion resolution goal.

One aim of the deferability concept is to have fewer cases where a congestion resolution maneuver was implemented but later turned out to be unnecessary. Delay recovery is much less frequent for the deferred resolution strategy because flights are not assigned delays or reroutes until shortly before departure. The downside is that NAS customers have less time to manage the impact of the maneuver, though they may be informed at an earlier decision point that the flight is likely to be maneuvered, if the (then deferrable) flight will potentially require a maneuver for congestion resolution.

5.2.9 Prediction Evolution

In order to capture the interesting features of probabilistic decision making, we must simulate how the state of knowledge (that is, the updated prediction) changes as simulation time passes. A single traffic and weather prediction exist at the start of the simulation. Many Monte Carlo outcomes are modeled from that prediction. When simulation time is advanced to the next decision point, each of those outcomes will also have an updated prediction, and that prediction will reflect what has become known since the last decision point. For example, if flight ABC123 is contained in the initial traffic prediction, but in a particular outcome ABC123 is cancelled, then there is some time at which this becomes known. If the flight is cancelled between the first and second decision points, then the prediction at the second decision point should not contain ABC123, and ensuing resolution actions will not attempt to delay or reroute that flight. Thus, we created models for how traffic and weather predictions ―learn" from observed events.

The traffic prediction evolution model is simple, but realistic enough to generate interesting results. Flights that cancel do so 15 minutes before their planned departure time. Pop-up flights file 30 minutes before their planned departure time. Flights that leave later than predicted are discovered to be late when their initial departure time passes. Flights that are rerouted receive the

new route at takeoff. These rules may be replaced in future with more realistic, statistically-modeled behavior based on empirical studies.

Weather predictions are evolved by "inverting" the Monte Carlo weather variation model. Recall that the sector coverage prediction error is captured in two parameters (speed and growth). These are generated from the triangular error distributions for each outcome. As time advances along the outcome, the speed and intensity errors become apparent, and the prediction is "reset" at each decision point. The coverage at that time becomes the new baseline, and new speed and intensity values are computed from the error distributions to represent the new prediction as variations from the new observation. The weather prediction thus learns from the experienced weather, though not perfectly.

5.2.10 Simulation Flow

Figure 5-7 illustrates the simulation flow. This assumes that the congestion resolution and management areas have been identified (CRA and CMA), baseline sector capacities are established, a risk management strategy has been chosen, and supporting data have been assembled (wind forecasts, Monte Carlo distribution parameters, etc.) The process begins with a predicted trajectory set and predicted weather coverage trace, which are used as a basis by the Monte Carlo traffic simulation to generate a set of N possible "actual" outcomes for the flights and sector capacities. These characterize the variety of ways that the situation can play out.

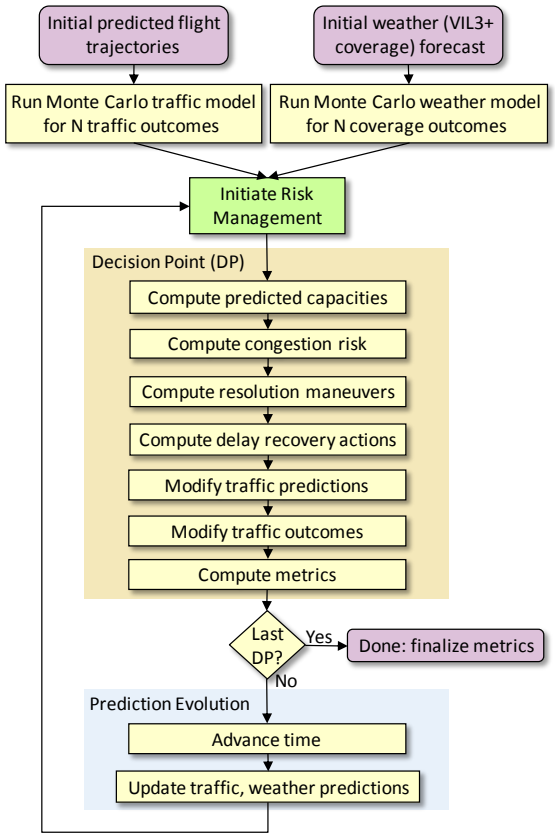

Figure 5-7. Simulation Flowchart

Next, the congestion risk management loop is started at the first decision point. First, the weather forecast is converted to sector capacity as described in Section 5.2.4. The congestion risk is computed by converting the predicted trajectories to sector traffic counts, applying the aggregate prediction uncertainty models (Section 5.2.5) to get traffic demand and capacity distributions, and convolving them. The resolution algorithm (Section 5.2.6) is run to determine the set of flight-specific maneuvers required to reduce predicted congestion to the congestion risk profile. No delay recovery is needed at the first decision point, as there are no prior maneuvers.

At this point, a new set of predicted flight trajectories is developed by substituting in the resolution maneuvers. The full set of actual outcomes must be updated to reflect the flight planning changes (route, departure time) that result from the resolution actions. The Monte Carlo traffic model is used again to adjust flight outcomes. If the resolution action for a flight includes a ground delay, then new ―actual" departure times are generated for that flight in each of the N traffic outcomes. If the resolution action includes a reroute, then the flight progress and route/altitude variability models are re-applied to that flight for all outcomes. This maintains consistency. Flights that are not maneuvered at a decision point retain the same set of trajectory variations at the next decision point, and the combined unmodified and modified set represents the altered range of traffic outcomes resulting from the executed resolution maneuver.

The last step in evaluating the decision point is to capture metrics. For both resolution and recovery, the number of flights affected, type and number of maneuvers generated, and the delays produced (or recovered) are saved. Predicted post-decision congestion probabilities are calculated to determine whether the resolution strategy succeeded. ―Actual" congestion probabilities, based on the modified Monte Carlo outcomes, are also saved.

The simulation now enters the prediction evolution step. Time is advanced to the next decision point, and the process described in Section 5.2.9 is used to update both the traffic demand and weather coverage predictions in each of the N outcomes to reflect knowledge gained with the passing of time. Note that at the first decision point, there was a single traffic demand prediction and a single weather coverage prediction. Now, since all N actual outcomes have different trajectory sets and different weather coverage traces, prediction updating produces different traffic demand and weather coverage prediction for each of the outcomes.

The simulation now loops back to do the decision point processing. For the second and subsequent decision points, the process is slightly different than for the first. Delay recovery is now possible, so after resolution maneuvers are calculated, the delay recovery algorithm described in Section 5.2.7 is used to identify any delay recovery actions that can be taken. The traffic predictions, and actual outcomes, must be updated with the results of both the resolution and recovery action sets to properly reflect the decision. Also, since there are now different predictions for each of the N outcomes, resolution and recovery actions must be calculated for each of the N outcomes independently and the results applied across all N outcomes.

The simulation continues until the last decision point is processed, at which time summary statistics are computed and the simulation is complete.

5.2.11 Implementation

The simulation has been developed in Java, and is highly parallelized. Because of the computational structure, groups of Monte Carlo outcomes can be independently carried through the decision sequence. Intermediate results are saved, and recombined for analysis after all parallel runs are complete. In the runs described below for the sample scenario, evaluation of each resolution strategy took approximately 2 hours to complete on 8 shared-use, dual-processor/dual-core systems (N = 1000). Certainly faster times would be possible using a dedicated cluster or grid computing environment.

5.3 Simulation Results for a Weather-Induced Congestion Problem

A realistic traffic scenario was developed by selecting traffic from a busy period of a clear-weather day, and creating congestion by creating an artificial moving weather coverage pattern which reduced capacities across the airspace. In this way we avoid using a real situation in which traffic management actions were taken to control congestion, making it hard to separate the effects of the method being tested from the actual traffic management activity. We ran the simulation for a variety of risk management strategies, for two levels of traffic management uncertainty, and for two different resolution action intervals.

5.3.1 Congestion Scenario

The area of interest for this scenario comprises four laterally or vertically adjacent sectors in ZDC. Three of the sectors (ZDC sectors 72, 16, and 36) are visible in Figure 5-8; the fourth, ZDC14, is a low altitude sector below 16 and 72. This area was designated as the CRA. The CMA, composed of all sectors adjoining the CRA either laterally or vertically, includes 38 sectors.

It is assumed that weather moves from west to east through the CRA such that coverage in each sector can be approximated as a half-sine wave (Figure 5-9). The solid curves indicate the weather coverage percentage as a function of time for the four CRA sectors. The corresponding sector capacities are shown by the dashed lines (referenced to right vertical axis). Each sector capacity drops from the nominal value to approximately half that value, and recovers about 3 hours later. It is assumed that the CMA is not impacted by weather, but the traffic levels are generally lower in those sectors and thus while the CMA could be considered, the results would not change much.

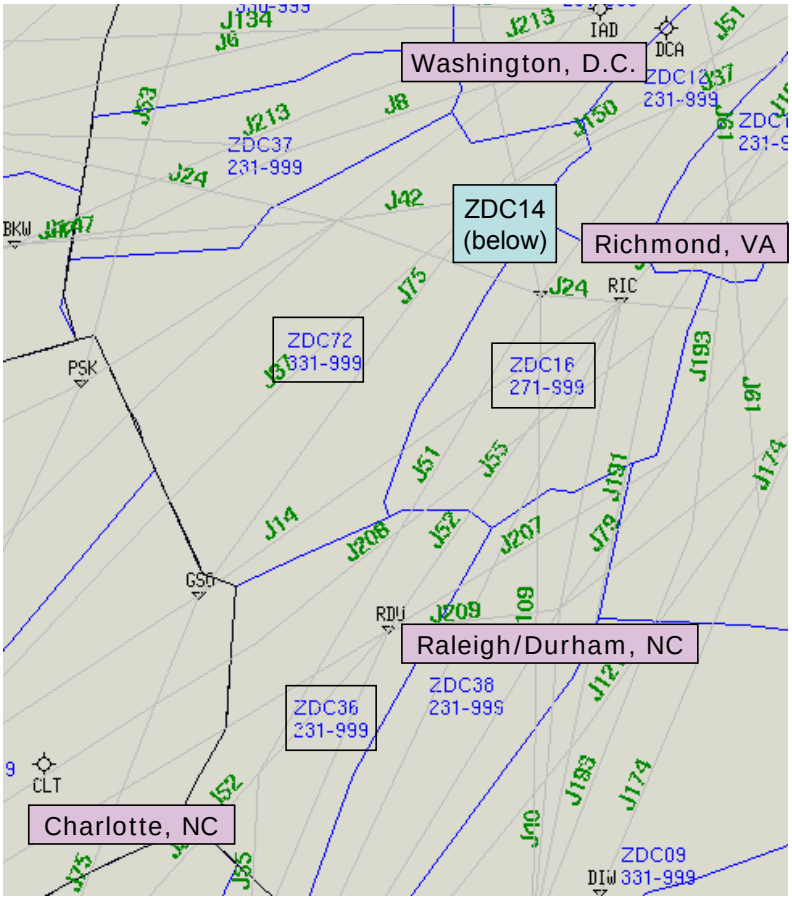

Figure 5-8. Congestion Scenario Airspace

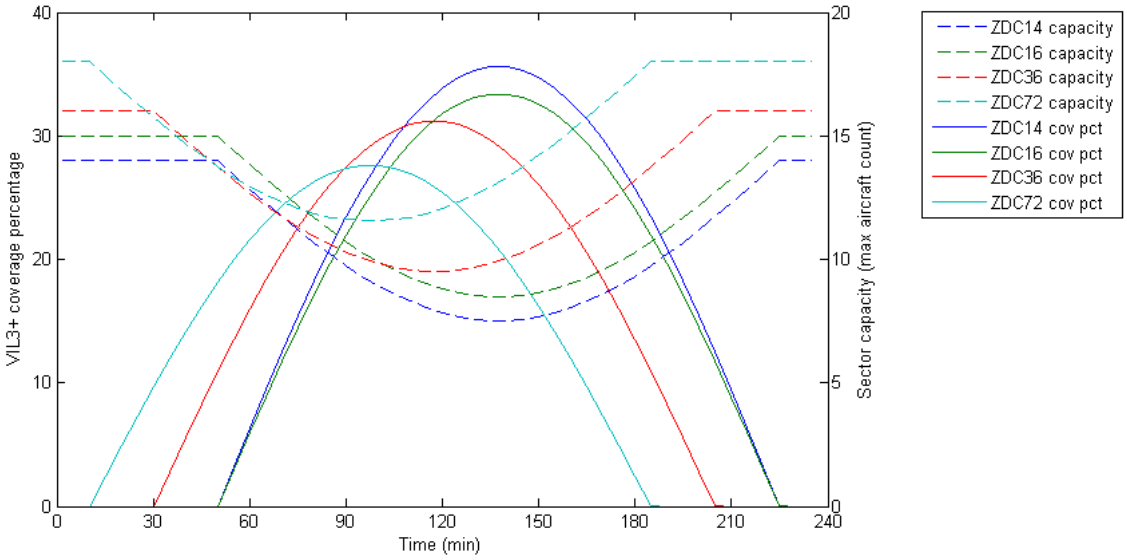

Figure 5-9. Predicted Weather Coverage and Resulting Sector Capacities at Start of Scenario

For this scenario, based on the nominal weather and traffic predictions, there is congestion predicted in the CRA sectors starting at 1715 UTC. Figure 5-10 shows the median peak traffic counts and congestion alerts. Each row of the matrix is a time-series prediction for one sector, at 15 minute intervals. The normal peak count threshold (MAP value) for each sector is next to the sector name. The number in each cell indicates the median peak traffic count value from the ADM. As in Figure 5-3, purple boxes indicate a greater than 75% probability that the actual demand exceeds the sector capacity. Pink boxes indicate a greater than 50% probability, and boxes with a gray background indicate a less than 50% probability. The majority of the anticipated congestion lies in the time period 1815 to 1914 UTC, and this represents a serious congestion situation that needs to be resolved. During this period, approximately 1500 flights pass through the CMA, of which about 200 pass through the CRA.

	1700	1715	1730	1745	1800	1815	1830	1845	1900	1915
ZDC14 [14]	9	12	8	6	6	9	10	12	11	10
ZDC16 [15]	7	7	7	10	9	10	11	11	5	5
ZDC36 [16]	6	10	8	14	9	13	15	9	6	7
ZDC72 [18]	16	19	12	8	15	15	18	19	13	13

Figure 5-10. Predicted Congestion at Start of Scenario

5.3.2 Risk Management Strategies

The strategies evaluated in this scenario are summarized in Table 5-1. In all cases, airborne rerouting was not allowed, and flights could be assigned maneuvers until 15 minutes before their departure time. Each strategy was run for a series of decision points starting at 1700 and ending at 1830, with a LAT of two hours, during which decision point intervals of 15 and 30 minutes were used. Thus, for the 15 minute interval, there were 7 decision points during the time period, and for the 30 minute interval, there were 4 decision points.

Table 5-1. Selected Risk Management Strategies

ID	0-15 min	15-30 min	30-45 min	45-60 min	60-75 min	75-90 min	90-105 min	105-120 min
RM1	0.5	0.6	0.65	0.7	0.75	0.8	1.0	1.0
RM2	0.5	0.5	0.6	0.6	0.7	0.7	0.8	0.8
RM3	0.5	0.5	0.5	0.5	1.0	1.0	1.0	1.0
RM4	0.5	0.5	0.5	0.5	0.6	0.6	0.7	0.7
RM5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
RM6*	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5

* denotes where the deferred resolution strategy (Section 5.2.8) was used

Table 5-1 describes the continual resolution strategies that were evaluated. RM1, RM2, and RM4 are varying-strength versions of the strategy found to work in the previous study, namely, increasing resolution risk targets with LAT. RM3 is a special case of a simplified strategy, which uses a risk goal of 0.5 out to 60 minutes, and simply ignores any congestion beyond that time.

RM5 is the "solve everything" out to 120 minutes strategy. RM6 has the same risk profile as RM5, but is effectively less aggressive because deferrable flights are not maneuvered immediately.

The resolution algorithm was asked to achieve the congestion risk profile by using ground delays of up to 30 minutes and/or by rerouting pre-departure flights on one of four alternate routes. These routes were chosen from an adapted route set which includes FAA-defined routes (for example, Coded Departure Routes) and a selection of historical routes that are frequently used. At least one of these routes was guaranteed to avoid the CRA (but not the CMA). No reroutes or ground delays were assigned to flights within 15 minutes of their planned departure time.

Delay recovery was also possible for all strategies, starting 60 minutes before the planned departure time. As described in Section 5.2.7, Parameter A was 15 minutes and Parameter B was 60 minutes.

We tested the 6 strategies in Table 5-1 under two different weather forecast uncertainty levels (moderate, high) and using two different decision point intervals (15 minutes, 30 minutes), yielding 24 total cases.

The statistical features of the output distributions indicated that 1000 Monte Carlo outcomes were required to obtain a 95% confidence that the estimate of the mean number of flights affected by each resolution action was within one flight of the actual mean.

5.3.3 Results

Table 5-2 shows the results of the 20 simulation runs in terms of mean metric values; being a stochastic simulation, all the metrics are distributions, and the nature of those distributions is important and will be discussed later. For both congestion resolution and delay recovery actions, there are three metrics related to impact on the NAS customers: number of flights affected by the solution, minutes of delay, and approximate direct operations cost (DOC) in dollars. Note that, for resolution actions, only positive delays are included so that credit is not given for rerouting a flight on a shorter path with an earlier arrival time. That situation is treated as "zero delay", since it is still a disruption, and NAS customers would generally like to fly their filed flight plan if possible. DOC is calculated from the assumption that airborne delays cost $50 per minute and ground delays cost $25 per minute.[18]

[18] Delay costs are subject to much variation (fuel prices, etc.) but these values are a reasonable approximation of costs in January 2009. In 2004, DOC was estimated by GRA, Inc. [GRA, 2004] for large aircraft as $43/minute for airborne delay and $23/minute for ground delay.

Table 5-2. Strategy Evaluation Metrics

Strategy	Interval (min)	Weather forecast unc.	Mean delay metrics			Mean recovery metrics			Means of other metrics			
			No. of flights	Delay (min)	DOC ($)	No. of flights	Delay (min)	DOC ($)	MTBD (min)	Delays/ affected flight	Recovery percent.	Unsolved congestion
RM1	30	Mod	80	381	15891	2	-8	-99	51	1.15	1.2	3
		High	81	443	17183	2	-9	-154	51	1.17	1.0	3
	15	Mod	96	466	19590	5	-26	-189	53	1.21	3.7	3
		High	101	558	21588	5	-28	-262	53	1.25	3.3	3
RM2	30	Mod	96	500	18976	3	-10	-122	55	1.18	1.6	2
		High	103	634	22266	3	-16	-270	53	1.20	1.4	2
	15	Mod	118	614	23477	8	-34	-257	57	1.24	4.5	2
		High	132	794	27876	9	-43	-534	56	1.32	4.0	2
RM3	30	Mod	80	364	15611	2	-4	80	47	1.10	2.2	3
		High	80	383	15831	2	-3	74	47	1.11	2.0	3
	15	Mod	102	482	20249	6	-20	39	49	1.19	4.8	2
		High	106	549	21685	6	-22	7	49	1.22	4.8	2
RM4	30	Mod	117	646	22514	4	-17	-246	55	1.18	1.8	2
		High	129	899	28593	5	-22	-375	55	1.23	1.5	2
	15	Mod	146	807	28094	13	-67	-857	58	1.27	5-4	2
		High	171	1186	37543	15	-81	-1281	59	1.40	4.9	2
RM5	30	Mod	171	1242	37367	45	-343	-6891	57	1.23	9.2	2
		High	189	1663	47857	47	-402	-8525	58	1.30	7.8	2
	15	Mod	219	1478	45602	94	-685	-13735	60	1.41	18.8	0
		High	260	2107	61705	101	-880	-18669	60	1.59	17.9	0
RM6	30	Mod	103	730	23200	3	-19	-415	39	1.12	0.7	2
		High	114	988	29671	3	-27	-625	39	1.15	0.5	2
	15	Mod	125	904	28529	8	-78	-1764	42	1.20	1.7	2
		High	146	1302	38737	9	-85	-1937	43	1.31	1.7	2

Three other metrics measure the characteristics of the resolution and recovery activity. Mean time before departure (MTBD) gives the average time before the original departure time that a flight received its first delay maneuver. Longer times are better for predictability. Delays-per-affected-flight indicates how often a flight that was maneuvered once was acted upon again. Since we allowed a maximum of 30 minutes of ground delay in a single action, a non-trivial number of flights needed to be affected more than once. The recovery percentage indicates what percentage of the affected flights received a full or partial delay recovery action.

Finally, the unsolved congestion column indicates how many CRA sector intervals (that is, cells in the matrix in Figure 5-10) had a congestion risk greater than 0.5 when the last decision point which could solve them was completed. For example, after the 1730 decision point with 30-minute decision intervals, CRA sectors for the 1730-1744 and 1745-1759 were examined, and any periods with congestion risk greater than 0.5 were added to the unsolved list. This is a simplistic way to evaluate the congestion effectiveness, because the value of the final risk is equally if not more important than the fact that it exceeds 0.5. A more detailed metric will be discussed later.

Some clear trends are evident in the results. More aggressive strategies delay more flights and incur more cost, but do a better job resolving congestion (more on this later). The mean cost of resolving the problem, for moderate weather forecast uncertainty, ranges from about $15,000 to $35,000 (if we sum the DOC means for delay and recovery actions). Also, more aggressive resolution action leads to more delay recovery actions, though this is not a strong effect until

very aggressive strategies are employed (RM4 with 15 minute interval, RM5 at either interval). More aggressive strategies are also more likely to maneuver a flight more than once (delays per affected flight).

Increasing the weather forecast uncertainty drives up the mean cost of solving the problem and slightly increases the need for delay recovery maneuvers. The weather forecast uncertainty has a larger effect on the variability of the measures, which will be discussed later. The ability to measure this effect has useful implications for testing the utility of new, probabilistic weather forecast products.

Running a given risk management strategy at tighter intervals (15 rather than 30) has the effect of increasing the strength of the strategy. This is seen in the table as increasing both the delay and recovery metrics, but it also increases congestion effectiveness.

The unsolved congestion periods metric works well enough to stratify the results into three groups color-coded in Table 5-2. Two of the sector periods were extremely difficult to solve, given the option space and prediction errors, and could only be solved with very aggressive action. Thus, only strategy RM5 with 15 minute intervals succeeded in doing so, at a very large cost. Most strategies solved everything but these two periods, except for the least aggressive strategy, which allowed another period to go unsolved, and the ―short" strategy (RM3), which left an additional period unsolved when run at 30 minute intervals. Thus, if we choose ―no more than 2 unsolved periods" as our congestion effectiveness target, then RM2 at 30 minute intervals is a good strategy for both moderate and high levels of weather forecast uncertainty. However, it is important to consider a variety of factors in deciding upon the best strategy, as discussed in Section 5.3.4.

The type of strategy chosen has a secondary effect on schedule variability, shown by the MTBD metric. Naturally, NAS customers find out later about maneuvers in solutions produced by RM6—the deferred resolution strategy—than they would find out about solutions produced by the other strategies. With the exception of RM3, which does not even attempt to resolve problems more than 60 minutes ahead of time, the other strategies all determine that flights will be maneuvered about 55 minutes before their planned departure time.

The deferred strategy RM6 has two other distinguishing characteristics. First, although the incurred cost is similar to that of RM4, RM6 affects significantly fewer flights. This will also be explored further in Section 5.3.4. Second, because maneuvers can be deferred, fewer delay recovery actions are needed in cases where the weather is not as severe as forecasted.

The more dramatic effect of weather forecast uncertainty is shown in Table 5-3. For all delay and recovery metrics, the standard deviation is much higher for the ―high" uncertainty level than for ―moderate," even though the mean value does not increase much. There is a secondary effect, which is that the more aggressive the strategy, the more sensitive it is to changes in the weather forecast uncertainty. The increased standard deviation means that the range of outcomes is greater, which has a more disruptive effect on airline schedules. The schedule disruption effect is not captured by the DOC metric, and may be an important consideration in choosing the best strategy.

Table 5-3. Behavior of Strategy Metric Distributions

Strategy	Interval (min)	Weather forecast uncertainty	Number of flights affected		Delay-induced DOC ($)		Recovered flights	
			Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
RM1	30	Mod	80	25.05	15891	5904	2	2.38
		High	81	33.94	17183	8703	2	2.61
	15	Mod	96	27.14	19590	7499	5	5.25
		High	101	40.92	21588	10671	5	5.53
RM2	30	Mod	96	26.66	18976	6980	3	3.12
		High	103	41.73	22266	12812	3	3.33
	15	Mod	118	28.95	23477	7734	8	6.65
		High	132	51.73	27876	14529	9	7.11
RM3	30	Mod	80	18.19	15611	5045	2	2.88
		High	80	24.02	15831	5947	2	2.81
	15	Mod	102	22.06	20249	6683	6	5.92
		High	106	30.22	21685	8036	6	6.07
RM4	30	Mod	117	28.65	22514	7979	4	3.99
		High	129	49.27	28593	16316	5	4.47
	15	Mod	146	33.83	28094	9271	13	9.42
		High	171	60.45	37543	18654	15	10.73
RM5	30	Mod	171	31.48	37367	10158	45	17.79
		High	189	43.68	47857	18725	47	22.93
	15	Mod	219	38.21	45602	11815	94	26.46
		High	260	55.01	61705	19850	101	34.59
RM6	30	Mod	103	29.64	23200	9218	3	2.37
		High	114	38.28	29671	15168	3	2.93
	15	Mod	125	32.01	28529	9940	8	5.05
		High	146	49.66	38737	19393	9	6.10

5.3.4 Defining Criteria for the "Best" Strategy

Summary statistics do not tell the whole story, since they obscure two important features of the solutions. First, they do not measure how severely capacity will be exceeded. Second, they do not show the variability in the results well, namely, how frequently across the outcomes sector capacities are exceeded, and what the range of possible cost values looks like. There are many ways to present this information, but we first need to decide what the best strategy looks like. This is an operational question, and we will seek input from operational traffic managers to determine the criteria. It might make sense to choose based on something like: ―choose the lowest cost strategy which reduces the probability of exceeding capacity by 3 flights to 15% or less."

Figure 5-11 illustrates the relationship between this type of goal, the incurred cost, and the weather forecasting uncertainty. The metric shown on the vertical axis represents the percentage of Monte Carlo outcomes in which one of the four CRA sectors has a peak count of 3 or more flights above its weather-impacted capacity during the 15 minute period following the last decision point in the simulation. This represents the effectiveness of the strategy in reducing the risk that significant congestion occurs at this point. For reference, this value is 59% if no resolution action is taken at all. This metric is plotted against the net mean cost (the resolution cost minus the recovered cost due to delay reduction) on the horizontal axis, and all 20 combinations of strategy, decision frequency, and weather forecast uncertainty are plotted. The red circles are for those combinations under high weather forecast uncertainty, fitted by a red regression line ($R^2 = 0.88$). Similarly, the blue triangles and blue fit line ($R^2 = 0.90$) represent these combinations under moderate uncertainty. Note that filled symbols denote strategy RM6, since it uses a different resolution method than the other five strategies.

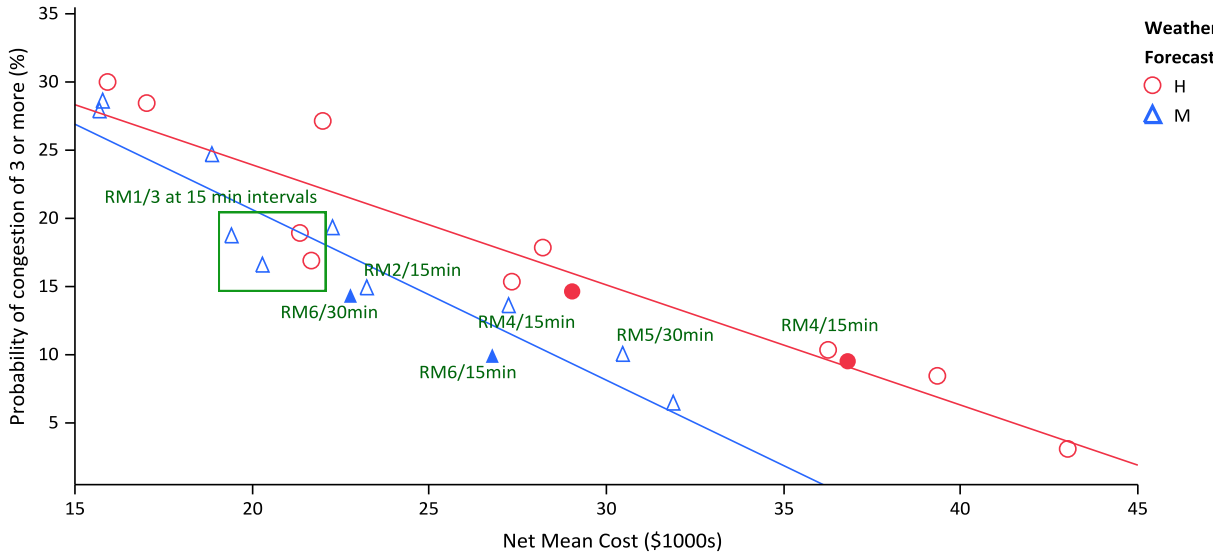

Figure 5-11. Congestion Risk Reduction Effectiveness versus Net Mean Cost

Several interesting features are revealed by this analysis. First, the difference between the linear fits indicates that under increased weather forecast uncertainty, it costs quite a bit more to reduce the significant congestion risk to a similar level. There is also wider variation in that cost, as discussed earlier. Also, the high quality of the fit implies that there is a relatively smooth linear relationship between the congestion risk and mean cost. The variations from this line indicate strategy-dependent differences. For example, the four points in the green rectangle are lower than the corresponding regression lines, indicating that they are more effective at reducing risk per dollar of mean cost. These four points represent strategies RM2 and RM3, executed at 15 minute decision intervals. This suggests that higher frequency decision-making is more cost-effective, though the severe congestion risk remains rather high for these cases.

It is also useful to examine a particular strategy under the two weather forecasts. Strategy RM4 succeeds in reducing the risk metric below 15% in both cases, which is logical since the

resolution algorithm accounts for the forecast accuracy in developing maneuvers. However, it costs 30% more under high weather forecast uncertainty. This is a little misleading, because the risk values achieved are not quite the same, due to statistical differences in the aggregate forecast error models. On average across the 10 strategy-frequency combinations, there is an 18% increase from moderate to high weather forecast uncertainty.

Note that the more aggressive strategies (hollow symbols at lower right in Figure 5-11) include not only delay actions, but a non-trivial number of delay recovery actions. Reducing the risk metric to very small values thus has an additional disruptive effect that is not captured by the net mean cost metric.

Finally, the effectiveness of the deferred resolution strategy (RM6) is evident, especially for moderate weather prediction uncertainty. The filled triangles show that RM6 incurs somewhat lower cost than other strategies to manage the congestion risk to a specific target level.

It is also useful to measure the number of flights affected by the various strategies, because it is another factor in evaluating schedule disruption. Figure 5-12 is a plot of the relationship between severe congestion risk and the mean number of affected flights, showing a similarly near-linear relationship as that for net mean cost. The same trends are observable, with the exception that RM6 performs even better with respect to this metric. Combined with the fewer delay recovery maneuvers required, it is clear that fewer maneuvers are required to achieve a target risk level using RM6, thus incurring less schedule disruption. This is somewhat offset by higher individual flight delays and shorter notification times than for other strategies.

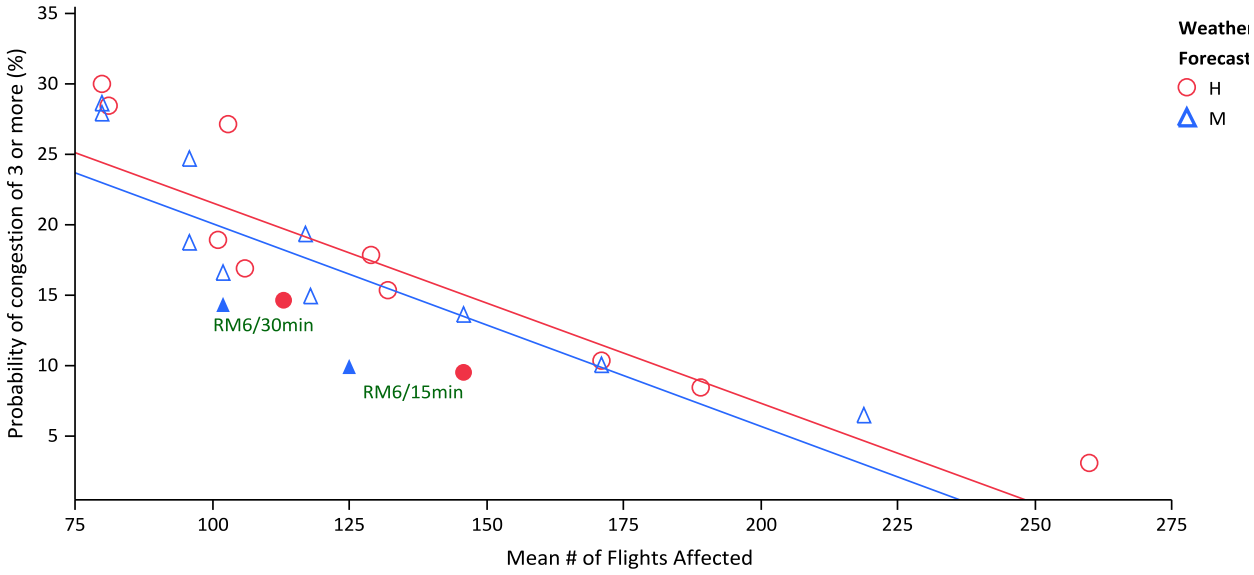

Figure 5-12. Congestion Risk Reduction Effectiveness versus Flights Affected

5.3.5 Variable Continual Congestion Resolution

In this study we used a time-invariant control strategy, in contrast to the decision-tree approach used in previous research [Wanke and Greenbaum, 2008]. The two methods could be combined

to get a better-performing continual resolution process. The decision to be made at each step would be between risk profiles rather than constant risk goals, that is, between less-aggressive and more-aggressive profiles. This decision could be made either by running a variant of this simulation in real time, or by running a large number of cases in offline simulations and developing heuristics which map common situations to effective strategies.

5.4 Applications

The method proposed here is suitable for real-time applications. Unlike the decision-tree approach, the Monte Carlo simulation is not needed for real-time operations, provided a suitable goal and risk management strategy can be chosen. We have, in fact, already implemented most of the method in a real-time decision support prototype. However, many issues still need to be addressed before an operationally-acceptable version could be deployed. These include: automatic updating of traffic and weather prediction probability models, cognitive engineering of the human-computer interface, incorporation of real probabilistic weather forecasts and a better weather-impacted sector capacity model, and how to best allow NAS customers to participate in resolution maneuver generation. The last could be handled by allowing NAS customers to submit preferred resolution options for their flights (already being discussed in government/industry working groups), or perhaps by automated negotiation between the resolution generator and airline flight planning software.

The results shown here are for a single, synthetic-weather scenario. More scenarios need to be analyzed to refine the strategies, and it may be that the variable-strategy version is needed to realize full benefits. Also, the resolution strategy developer we used can be improved. Even with the simple algorithm shown here, there are variations that might improve the utility of the approach. Finally, the one-pass maneuver assignment approach is far from optimal, and we have done work on genetic algorithms [Sood et al., 2007] and a partial-optimization procedure (described in Section 6) to gain improved solutions.

A second use for the method and simulation is cost-benefit analysis. It is possible, for example, to quantify the effect of weather forecast uncertainty on decision making and cost. Thus, the simulation is a platform for evaluating the potential tactical congestion management benefits from proposed probabilistic weather forecasting products, provided an operationally-acceptable model for weather-impacted sector capacity can be developed. Also, if a new technology is proposed that reduces uncertainty in demand or capacity prediction (for example, a surface management system, which would reduce departure prediction uncertainty), then the delay reduction benefits can be estimated via simulation.

5.5 Summary and Next Steps

A practical real-time method for probabilistic tactical congestion management has been developed. Simulation results indicate that a deferred resolution strategy using congestion risk targets that gradually increase with look-ahead time works better than the other strategies tested, but more simulations are needed to determine the precise parameters that work the best across a range of congestion situations. The effect of weather forecast uncertainty is clearly evident in the

results, and when real probabilistic forecasts are available, it will be interesting to identify the impact of these on congestion resolution effectiveness and cost.

There are a number of possible research directions for this work. In FY10, we will focus on two areas. The first is to establish a concept of use for the capability described here, as a step towards developing functional requirements. This will involve fleshing out some of the details we have assumed away in the analysis described above, such as integration with real weather products, the precise method of collaborating with NAS customers, and testing with additional congestion scenarios to settle on an initial risk management strategy.

The second area of immediate interest relates to how automated, tactical, flight-specific congestion management fits into the larger picture of operational flow management. Given the uncertainties when predicting weather and traffic at LATs longer than 2 or 3 hours, it is quite difficult to manage specific flight trajectories. It is likely that aggregate flow management methods will continue to be used for strategic decision making, and that different methods should be employed to do probabilistic planning. Thus, we plan to study the integration and interactions between aggregate-level strategic and flight-specific tactical decision making.

6 Heuristic Optimization for Developing Congestion Resolution Maneuvers

The congestion resolution algorithm described in Section 5.2.6, referred to as a "greedy heuristic" method, is a simple, fast, but non-optimal way to develop resolution maneuvers in a real-time or fast-time context. Heuristic optimization methods can be applied to complex problems like this one to improve resolution performance (that is, incur lower delays for the same congestion management effectiveness). In previous work, a genetic algorithm [Mulgund et al., 2006] and a multi-objective hybrid genetic algorithm [Sood et al., 2007] showed potentially significant improvements in solution quality over the non-optimized solution, but at a great computational expense. Therefore, a better heuristic which provides good solutions at low computational expense is desirable.

An alternative heuristic optimization approach is the Generalized Random Adaptive Search Procedure (GRASP), applied as a modification of the existing greedy heuristic procedure. GRASP introduces stochastic elements into an otherwise deterministic solution process to potentially provide better solutions without incurring the computational expense of more traditional heuristic optimization approaches. The GRASP algorithm has been shown to aid in problems in which an ordering scheme must be defined but the exact nature of a desirable prioritization criteria is not fully understood [Feo and Resende, 1995; Taylor et al., 2007].

The implementation of GRASP for FY08 research [Wanke et al., 2008a; Taylor and Wanke, 2008] examined the impact of multiple prioritization criteria on a set of single-objective metrics and in almost every instance, the GRASP results showed a statistically significant improvement over a deterministic greedy baseline. This research further explores the utility of the GRASP approach through a detailed examination of the prioritization criteria in order to define a more desirable heuristic rule. Having identified an alternative prioritization criterion, both single-metric and multi-metric objectives are constructed to understand the benefits and costs of valuing different goals.

6.1 GRASP Algorithm

The GRASP algorithm evaluates multiple potential solutions and returns the best solution found. Potential solutions are defined by randomly perturbing the prioritization scheme and resulting sort order of the candidate flight list, where the sorted candidate flight list determines which flights receive an undesirable resolution action in a given problem. GRASP was utilized to obtain benefits over the deterministic greedy heuristic procedure (Section 5.2.6), but operates under a similar framework.

Figure 6-1 shows the overall implementation of the GRASP algorithm. As we can see from Figure 6-1, GRASP develops a sorted candidate flight list that modifies the absolute implementation of a priority order to account for unknown probabilistic impacts. Given that the sorted list is generated probabilistically, each iteration can produce a slightly different list which can change the overall impact on the system. Therefore, multiple iterations are performed until

the termination criteria are met. A key aspect of the GRASP algorithm is how prioritization criteria are used to formulate the sorted candidate flight list.

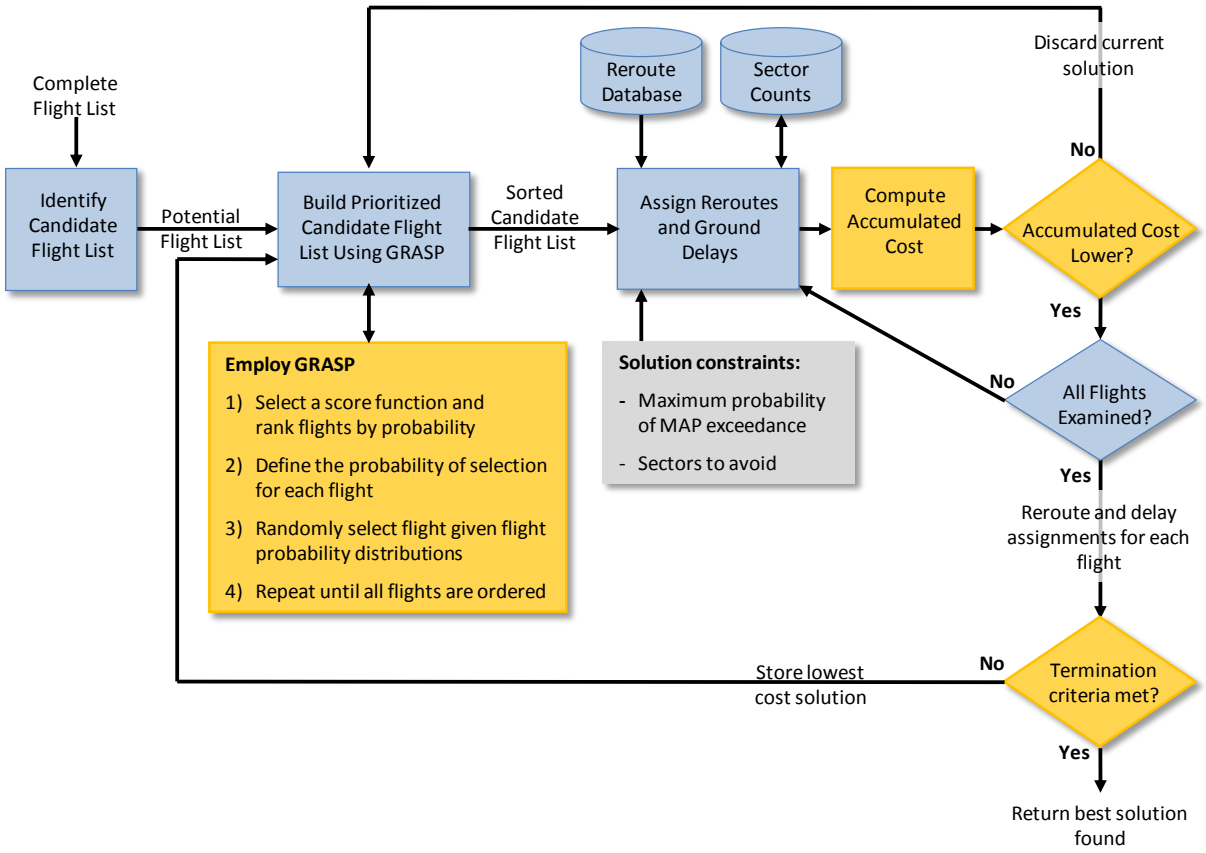

Figure 6-1. Overview of GRASP-based Congestion Management Algorithm

6.1.1 Prioritization Criteria Selection

The GRASP algorithm uses prioritization criteria to influence the position of a flight in the sort order, in which priority in the sort order reflects a decreased likelihood of being assigned an undesirable resolution action. Therefore, it is desirable that the prioritization criterion selected align with the objective function, so that the ordering of the flights promotes good solutions.

This research considers two different prioritization criteria, namely time to CRA and total time spent in the CRA.[19] Using the prioritization criterion of time to CRA prioritizes flights entering

[19] Previous research considered an additional prioritization criterion, but based on previous findings and current research objectives, these two prioritization criteria were chosen. However, other possible prioritization criteria, or a combination of multiple criteria, may be desirable to promote more effective sort orders.

the CRA sooner than other flights. This criterion represents how flights are traditionally prioritized in order to promote transparency and fairness in the decision process. Prioritizing by time spent in the CRA reflects that some flights have a greater impact on the system because they travel through the CRA longer. Therefore, if the flights that are in the CRA longer are prioritized lower and receive a resolution action, that resolution action may have a larger marginal benefit to the system than a resolution action for another flight that only minimally impacts the system. For each flight, the prioritization criteria is computed based on the original route of the flight and scaled to appropriate units as necessary.

6.1.2 Development of a Sorted Candidate Flight List

A key aspect of the GRASP algorithm is in the construction of the sorted candidate flight list. Instead of strictly ordering the flights by the prioritization criteria, the GRASP algorithm utilizes a set of score functions to inform the sorted flight list construction. The score functions represent different weightings of the prioritization criteria and are constructed such that lower scores are more beneficial than higher scores for priority in the sort list.

For each prioritization criteria (pc) selected, multiple score functions are defined that vary the relative weighting of priority importance between flights. The score functions used to evaluate the prioritization criteria described above are:

$$S_1 = pc$$
$$S_2 = \sqrt{pc}$$
$$S_3 = pc^2$$
$$S_4 = pc^{1.5}$$
$$S_5 = \sqrt[3]{pc}$$

In each iteration of GRASP, one of the score functions is selected uniformly at random from the set of defined score functions. Using the selected score function, a score is calculated for each flight in the candidate list; however this score does not automatically determine the ordering of the flight in the sorted candidate list. Instead, the GRASP algorithm uses these scores to determine the *probability* of selection of a given flight for priority ordering. That probability of selection is defined as

$$\boldsymbol{P(f) = \frac{e^{-s_i^f}}{\sum_f e^{-s_i^f}}} \qquad \text{Equation 6-1}$$

in which f is a flight in the candidate list, S_i^f is the score of flight f evaluated for score function i, and $P(f)$ is the corresponding probability of selecting that flight. As the summed probability over all flights is one, the probability of each flight can be viewed as the relative proportion of selection.

Figure 6-2 illustrates this process through a simplified example. Each sector of the circle represents the probability of selection of a given flight. For each slot in the sorted candidate

flight list, the wheel is —spun" and the next flight is selected. These probabilities are computed by defining the prioritization criterion and GRASP randomly selecting a score function.

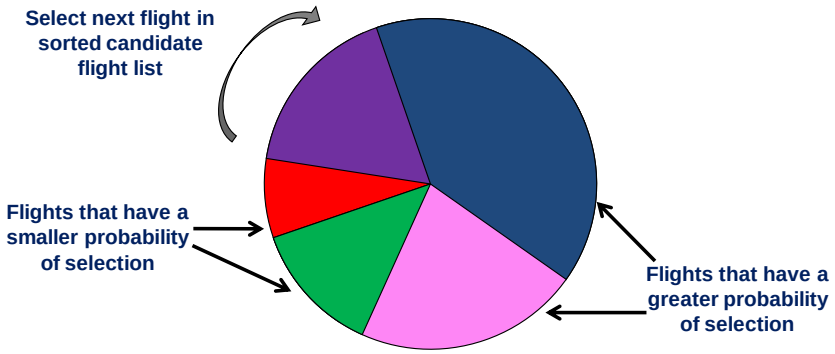

Figure 6-2. Diagram of Example Flight Probability Distribution

Flights with high probabilities of selection are more likely to be selected early in the process and therefore have a high priority on the sort list. However, the relative probability values do not determine the final sort order in the flight list. Instead, a flight is selected when a randomly generated number falls within the probability sector of the flight. This process is repeated until all flights have been ordered in the candidate list.

6.1.3 Measuring Solution Quality

Multiple metrics are available to assess the quality of the resolution strategies developed, depending on organizational priorities. For this research, three metrics were defined: total delay, probability of congestion, and inequity. The goal is to minimize the value of the objective function, defined as one or more of those metrics. Under all metrics, the congestion target is provided and resolution maneuvers are selected to best meet this target; however this target cannot always be met given previous decisions and the desire to meet other objectives.

The metric of total delay (J_d) selects the solution that provides the minimal increase in total system delay while attempting to meet the congestion target. Both ground delay (g) and airborne delay (a) contribute to the total delay. Only positive airborne delay is considered here to ensure that resolution actions that reroute flights to arrive at their destinations earlier than originally planned are neither penalized nor rewarded. The mathematical expression of delay is the sum of those delays:

$$J_d = \sum_f [g_f + \max(a_f, 0)] \qquad \text{Equation 6-2}$$

The metric of congestion compares potential solutions on the basis of which option produces the minimum sector congestion. The congestion metric (J_c) reflects the system level goal of reducing sector congestion, in which the impact on individual flights is a secondary consideration:

$$J_c = \sum_{i=1}^{N_c} \left[(\lambda_a (P_i - P_m) + \lambda_b (P_i - P_m)^2) \lambda_c \left(\frac{L_m - L_i}{L_m} \right) \right] \quad \text{Equation 6-3}$$

The quadratic cost term for positive deviations of congestion from the target level gives extra weight when the congestion of a sector (P_i) deviates from the sector specific target level (P_m). A decreasing linear cost on look-ahead time is included to represent that congestion occurring at a given LAT (L_i) is less concerning, all things being equal, when it is closest to the maximum look-ahead time (L_m). The weighting factors λ_a, λ_b, and λ_c represent different relative weightings of the components in the metric, in which the value of these factors were previously chosen through an off-line experimental study [Sood et al., 2007]. The total congestion is added for every sector and look-ahead time combination (N_c) in which congestion is present.

The final metric considered is inequity. Equity, as defined in this research, is when the delay incurred by the assignment of resolution maneuvers is as evenly distributed as possible over the different NAS customers. Therefore, the inequity metric (J_e) is defined as the minimum standard deviation of delay between NAS customers. The total number of flights affected by delay is represented by N_F and the number of customers with these affected flights is represented by N_{ar}. For the purpose of this analysis, all general aviation flights are grouped together as a single customer.

$$J_e = \frac{\sum_{k=1}^{N_{ar}} \left(\overline{Delay}_k - Delay_{N_{ar}} \right)^2}{N_{ar} - 1} \quad \text{Equation 6-4}$$

in which

$$\overline{Delay}_k = \frac{\sum_{f=1}^{N_F} [g_f + \max(a_f, 0)]}{N_F} \quad \text{Equation 6-5}$$

$$Delay_{N_{ar}} = \frac{\overline{Delay}_k}{N_{ar}} \quad \text{Equation 6-6}$$

6.1.4 Assigning Resolution Actions

Given the prioritized candidate list, the algorithm proceeds by selecting each flight in turn and evaluating its nominal path against the congestion risk target to determine what (if any) action needs to be taken. The flight is then assigned a route and the total system cost incurred for that flight is computed, based on the selected metric to be minimized.

Because the GRASP algorithm minimizes the total system cost, the cost incurred by the resolution action defined for each flight is iteratively added to the system cost, in order to measure the performance of the current solution. If the current solution's accumulated cost is greater than the minimum cost solution found up to that point, the iteration is terminated and a new priority list is constructed. If instead, after all flights have been evaluated, the current solution has less total cost, it replaces the best solution found. The best solution found over all iterations is returned when the termination criteria, in this case defined to be the maximum number of iterations, is met.

6.1.5 Example Congestion Problem

In order to evaluate the performance of the GRASP algorithm, we constructed an example and compared the results provided by each solution method to determine how effectively they resolve the congestion. The example congestion problem was developed from traffic patterns and predictions observed during January 2004. The traffic predictions used were from a TFM decision-support prototype [Wanke et al., 2003] that uses a probabilistic traffic demand model, namely the first version of our ADM described in [Zobell et al., 2005; Wanke et al., 2005a], to determine the probability that the demand on a specific sector exceeds the normal MAP value.

Figure 6-3 illustrates the geometry of the airspace structure around ATL, where the example problem takes place. The CRA consists of two sectors: ZTL38 and ZTL39. ZTL38 is a low altitude sector, controlling altitudes from 10,000 up to 24,000 feet, that primarily handles the departure traffic over fixes NOONE and NOTWO. Departures from NOTWO transition into sector ZTL39, which is a high altitude sector handling traffic from 24,000 up to 35,000 feet. ZTL39 also handles a complex pattern of overflight traffic, as illustrated by the jet airways intersecting at the VXV navigational aid.

Figure 6-4 shows the congestion situation in the form of predicted sector counts. This is a probabilistic Center Monitor that uses the ADM to estimate peak traffic counts and generate alerts. In this version, the median predicted peak count for each sector is shown, thus compensating for prediction biases. Also, the purple and pink alert colors are based directly on probabilities of congestion. A purple alert indicates a higher than 75% probability that the actual peak traffic demand will exceed the MAP. A pink alert indicates a 50% to 75% probability. The maximum congestion target for the example problem defined is 0.5, meaning that the goal is to obtain a solution that reduces congestion probabilities below 50%.

In Figure 6-4 both sectors 38 and 39 show congestion alerts for the time periods with 45 to 90 minutes LAT (between 1445 and 1530), with a maximum probability of congestion of 0.96. Flows through these two CRA sectors were analyzed to determine all possible sectors to which flights in the CRA can exit, and from which sectors flights can enter. Those 29 adjacent sectors constitute the CMA for this problem and have a maximum probability of congestion of 61%.

The goal of the GRASP algorithm is to determine a solution that meets the congestion target of 50% in the CRA, without adversely impacting the congestion in the CMA and with minimum total cost. For each prioritization criterion and objective function defined, the GRASP algorithm was run twenty-five times, with each run consisting of 100 iterations, and the results presented are averaged over all runs.

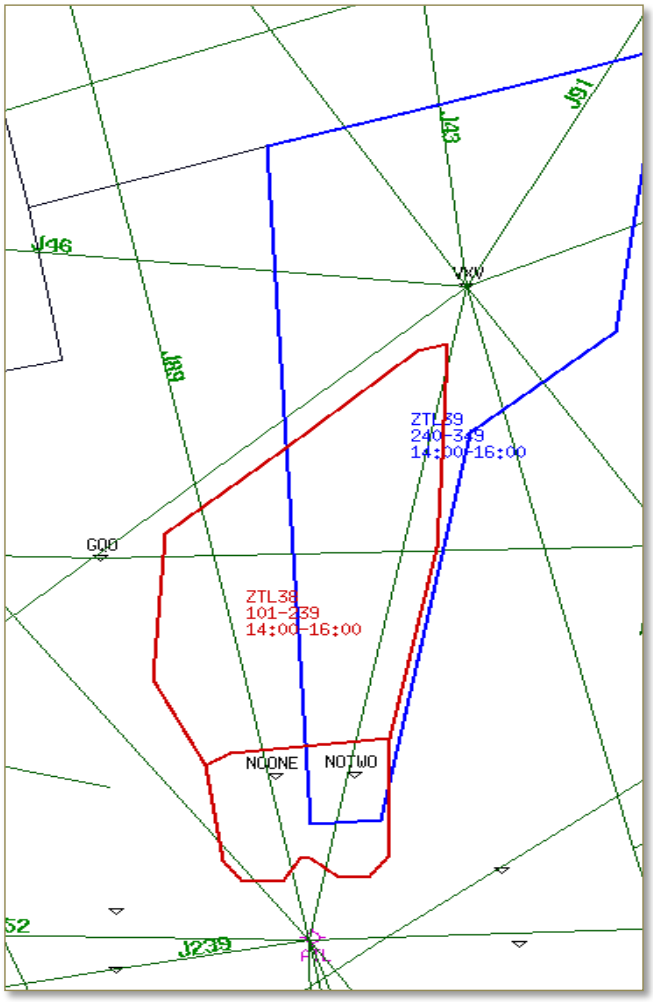

Figure 6-3. Geometry of Example Problem

	1400	1415	1430	1445	1500	1515	1530	1545
ZTL34 [12]	0*	2	4	4	8	9	6	2
ZTL36 [13]	0*	2	3	4	8	9	9	3
ZTL37 [13]	1	2	2	7	10	12	13	4
ZTL38 [13]	1	5	7	14	16	14	10	2
ZTL39 [13]	1	2	11	8	20	21	10	6
ZTL40 [18]	1	4	7	6	6	8	9	5

Figure 6-4. Probabilistic Congestion Display for Example Problem

6.2 Prioritization Criteria Selection Manipulation

In FY08 we showed that the two prioritization criteria, time to CRA and time spent in the CRA, provided improved solution quality as measured by the three metrics considered. However, there was not a single prioritization criterion that provided high quality solutions for each metric. Instead, the choice of prioritization criteria varied. Consequently, FY09 research delved deeper into the underlying dynamics of the prioritization criteria and score function selection to develop alternative criteria.

6.2.1 Impact of Score Functions

The value of implementing a variety of score functions and randomly choosing a score function for an iteration lies in the change of the relative probability of selection in the sorted order for a given flight. Modifying the relative probability of selection in effect changes how likely flights are to be swapped in the sorted list order, potentially leading to improved solutions.

The score functions described in Section 6.1.2 provide five different probability distributions for each of the two prioritization criteria examined. The probability distributions are computed for each prioritization criterion using the original flight information of the 64 flights in the example problem described in Section 6.1.5. The curves are generated by applying a fit to the histogram of probabilities generated using the given prioritization criterion and the chosen score function.

Figure 6-5 shows the probability distributions associated with the score functions when selecting the prioritization criteria of time to CRA. Examining Figure 6-5 shows that each score function produces a different probability distribution; however they all do share the general characteristic that a high proportion of flights have a probability near 0.015. For example, taking the distribution of score function 5 (S_5), we see a high peak at around 0.015. Alternatively, examining score function 3 (S_3), we see a smaller peak at around the same probability; however instead of falling off as drastically at higher probabilities, a smaller but significant number of flights have a probability of 0.02. The probable impact of these different probability distributions in determining the sort order is as follows.

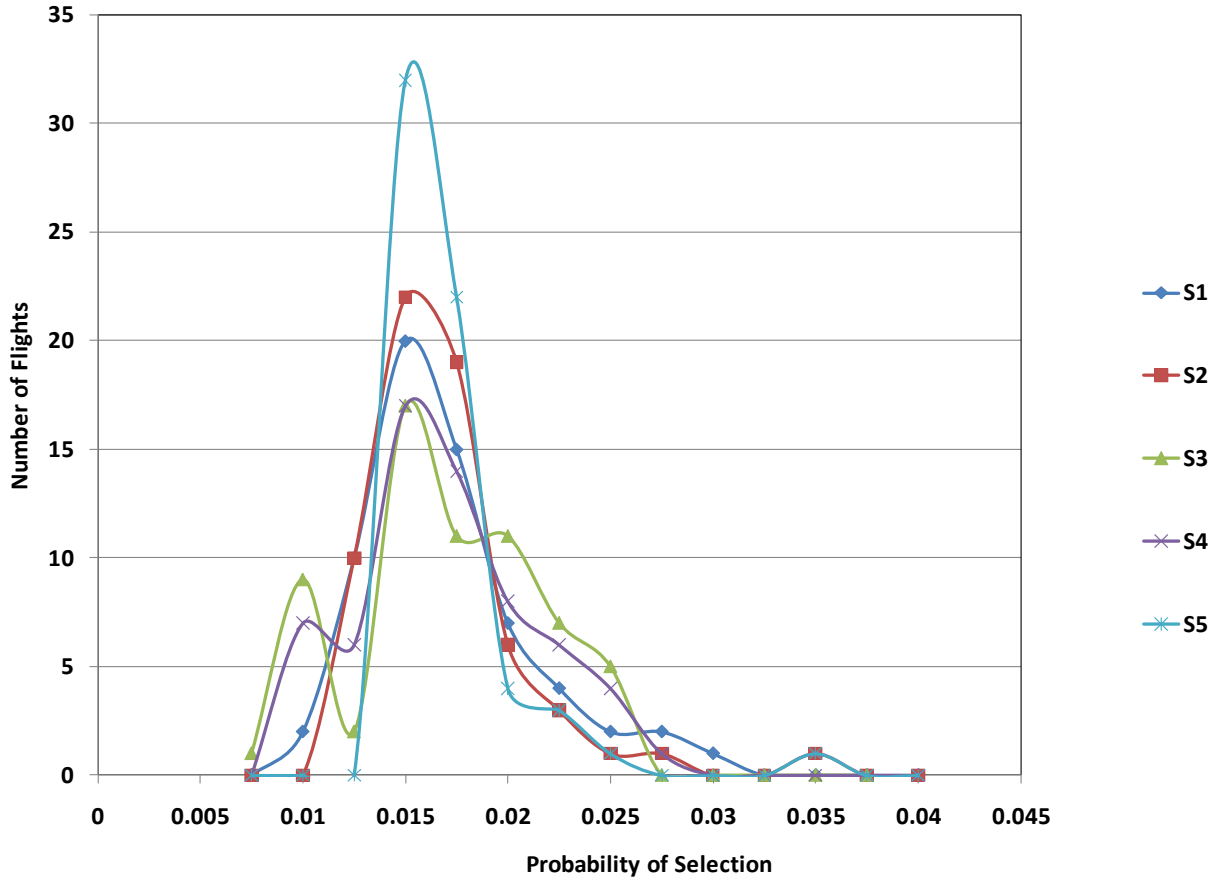

Figure 6-5. Probability Distribution of Score Functions for Time to CRA

Using score function 5 a few flights have high probabilities of selection and therefore will most likely be selected first in the sort order. The majority of flights has relatively the same probability of selection and therefore can easily be interchanged, thereby producing vastly different sorted lists. Alternatively, using score function 3, there is a larger selection of flights with higher probabilities, yielding a greater interchange of orders at the top of the list. These flights, however, are more likely to be chosen before the flights with lower probabilities. Effectively, this probability distribution bins the flights in the sort order, allowing only localized movement in the list.

Examining the probability distributions for the score function when using the prioritization criteria of time spent in the CRA reveals a different picture. Figure 6-6 shows a more uniform probability distribution for each of the five score functions, where the general shape is the same for all, with only small shifts in height, width and location of the peaks. Selecting the prioritization criterion to be the time spent in the CRA produces different ordering than time to the CRA; however, both have been shown in previous research [Taylor and Wanke, 2008] to improve the solution quality as compared to solutions produced through deterministic orderings, but more improvements can be gained. Furthermore, we are searching for a single prioritization criterion that consistently produces high quality solutions under multiple objective functions.

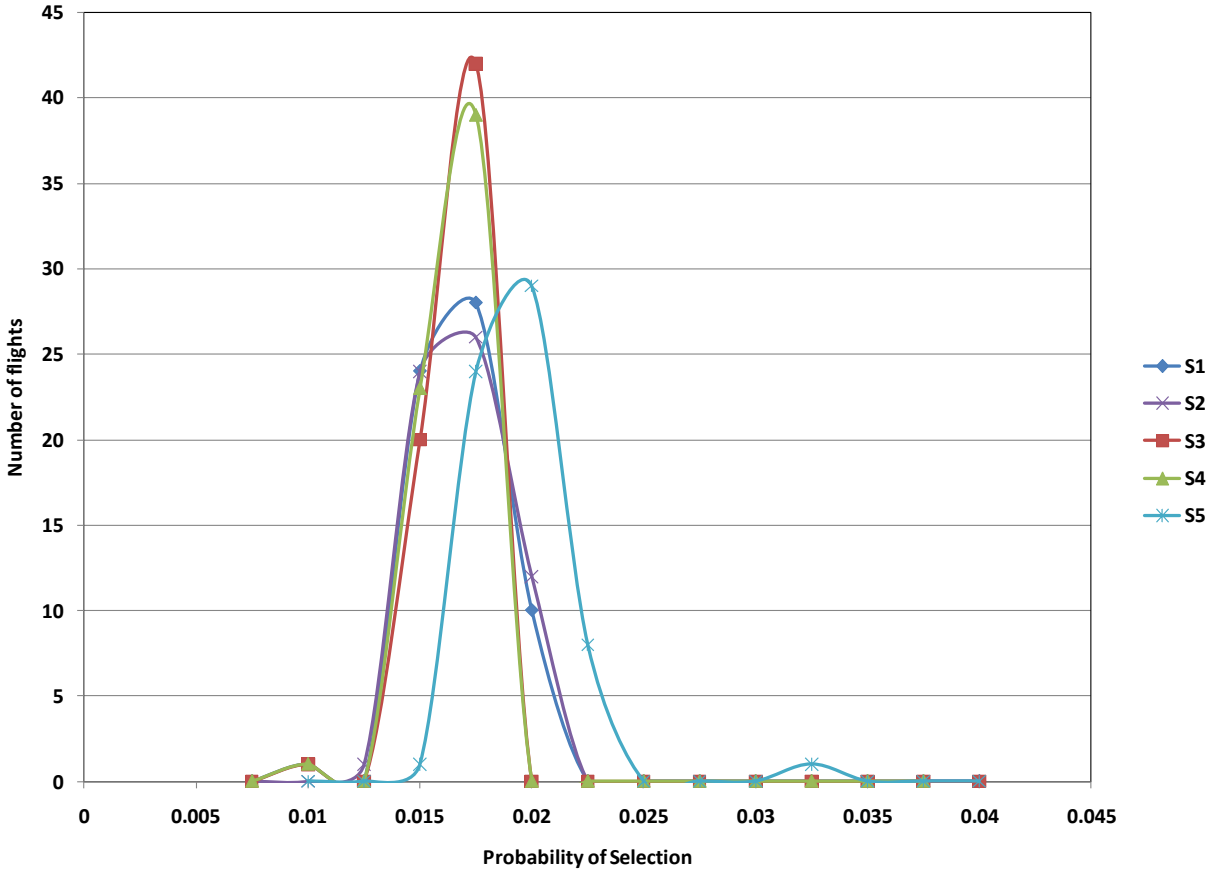

Figure 6-6. Probability Distributions of Score Functions for Time Spent in the CRA

6.2.2 Combined Prioritization Criteria

Given the resulting probability distributions for the original score functions examined under each of the prioritization criteria, new score functions were defined that vary the weighting of the two prioritization criteria together when developing the sorted lists. The new score functions considered are:

$$S_{1*} = pc_1 * pc_2$$

$$S_{2*} = \sqrt{pc_1} * pc_2$$

$$S_{3*} = pc_1 * \sqrt{pc_2}$$

$$S_{4*} = \sqrt{pc_1} * \sqrt{pc_2}$$

$$S_{5*} = pc_1^2 * pc_2$$

$$S_{6*} = pc_1 * pc_2^2$$

where $pc_1$is the time to CRA and pc_2 is the time spent in the CRA. These six score functions were chosen out of the many potential combinations of the two prioritization criteria to provide a

variety of different probability distributions. Figure 6-7 provides the probability distributions for the new score functions. Examining Figure 6-7 shows how the score functions produce significantly different distributions, but are related to the distributions shown in Figure 6-5 and Figure 6-6.

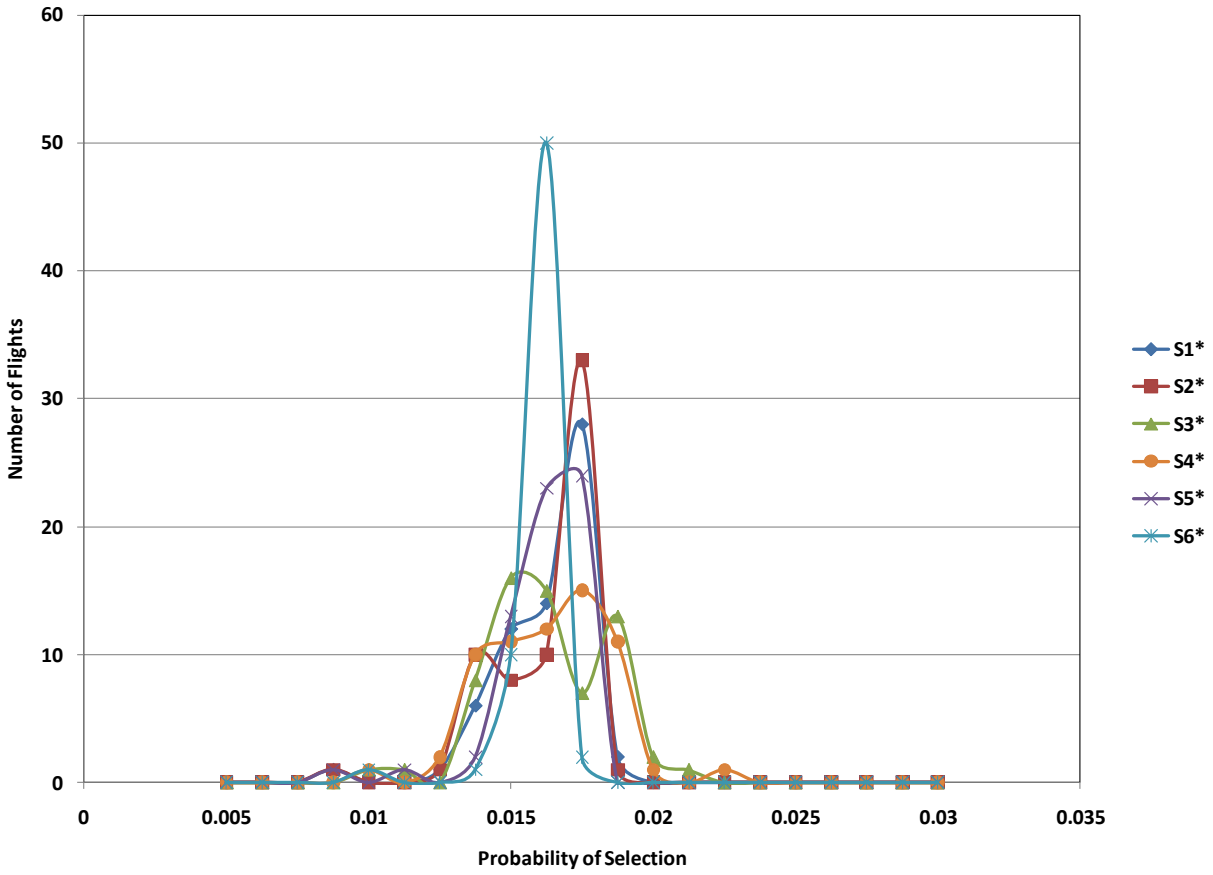

Figure 6-7. Probability Distributions of Score Functions for Combined Prioritization Criteria

The six new score functions were implemented in GRASP for the single-metric objectives of minimum delay, congestion, and inequity. The results were compared to the metric values produced by employing the original prioritization criteria using the original score functions. Again, the results presented are the averaged results for 25 runs of 100 iterations each. For clarity, future references to prioritization criteria and score functions will follow these conventions:

- "Time to CRA" refers to using the prioritization criterion of time to CRA with the original 5 score functions.
- "Time Spent in CRA" refers to using the prioritization criterion of time spent in the CRA with the original 5 score functions.

- ~~"C~~Combined prioritization" refers to using both the time to the CRA and time spent in the CRA as the prioritization criteria in the new 6 score functions.

For the objective of minimum delay, Figure 6-8 shows the impact on each of the three metrics. Although only delay is minimized, all metrics are evaluated. Examining Figure 6-8 shows that ~~"c~~ombined prioritization" provides the lowest delay and the median valuation point for congestion and inequity. This improves upon the previous findings where ~~"ti~~me spent in CRA" provided lower delay but much higher congestion and inequity than ~~"ti~~me to CRA." Thus, for minimum delay, the combined prioritization may provide an improved prioritization criterion.

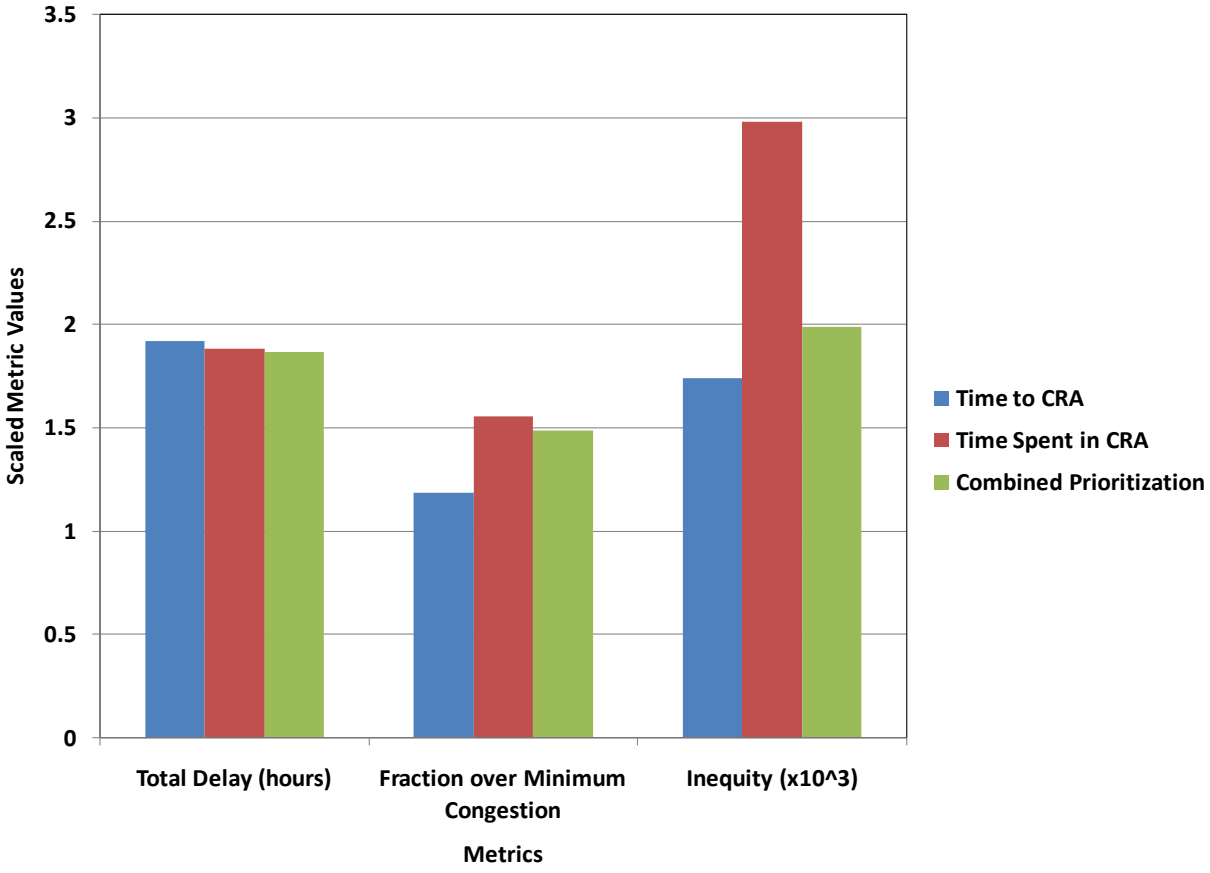

Figure 6-8. Prioritization Criteria Comparison for Minimum Delay

For the objective of minimum congestion, Figure 6-9 shows the impact on each of the three metrics from implementing the three prioritization criteria. Examining Figure 6-9 shows that the ~~"c~~ombined prioritization" provides the same minimum congestion as the other two prioritization criteria. Additionally, it provides the median valuation in the other two metrics. Unlike the minimum delay condition, here ~~"ti~~me spent in CRA" produces the minimum congestion and the lowest delay and inequity, thereby decreasing the necessity of finding an alternative prioritization criterion for this objective function.

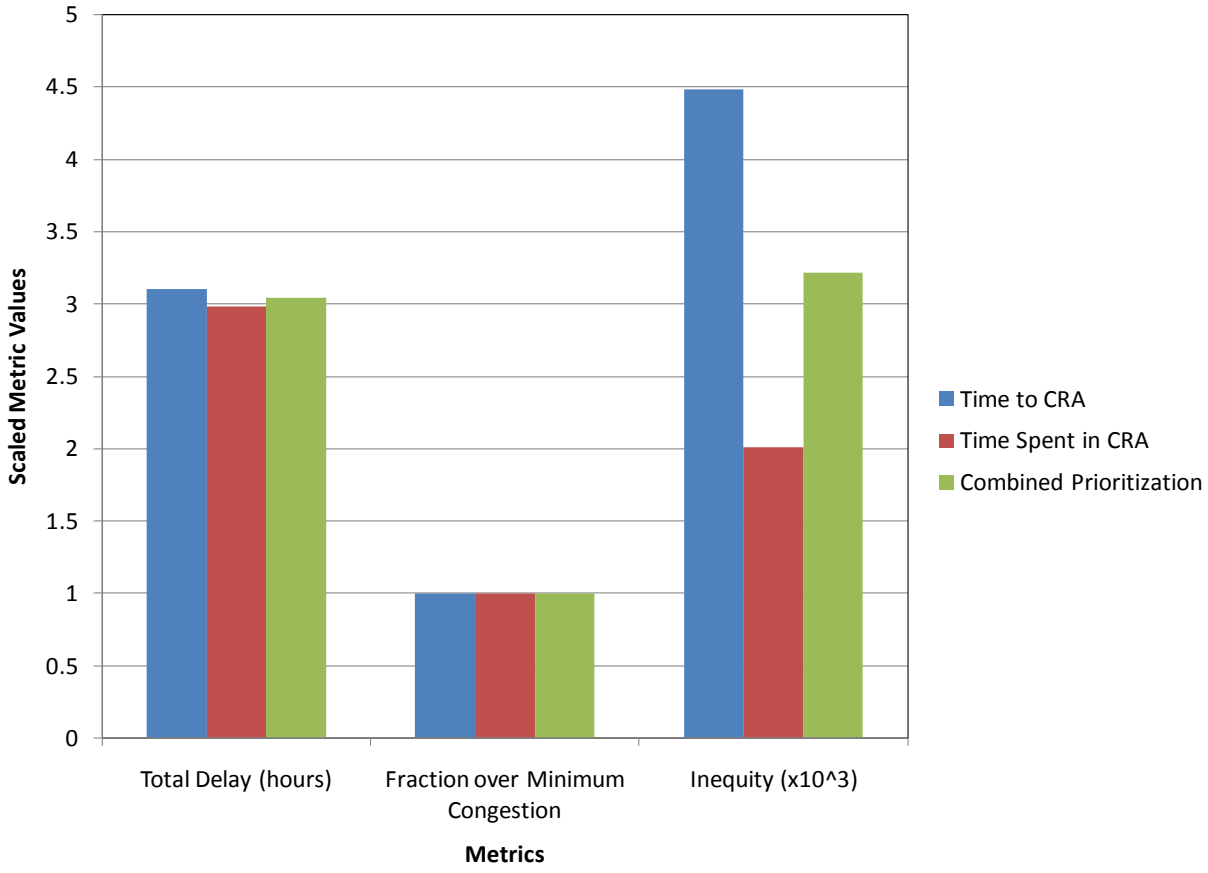

Figure 6-9. Prioritization Criteria Comparison for Minimum Congestion

For the objective of minimum inequity, Figure 6-10 shows that all three prioritization criteria provide an extremely low inequity value, with "time to CRA" providing the lowest value. "Time to CRA" also provides the lowest congestion but the highest delay. The lowest delay comes from "time spent in CRA." The "combined prioritization" provides either the median or high value in each metric, and therefore may not be a desirable criterion for this objective.

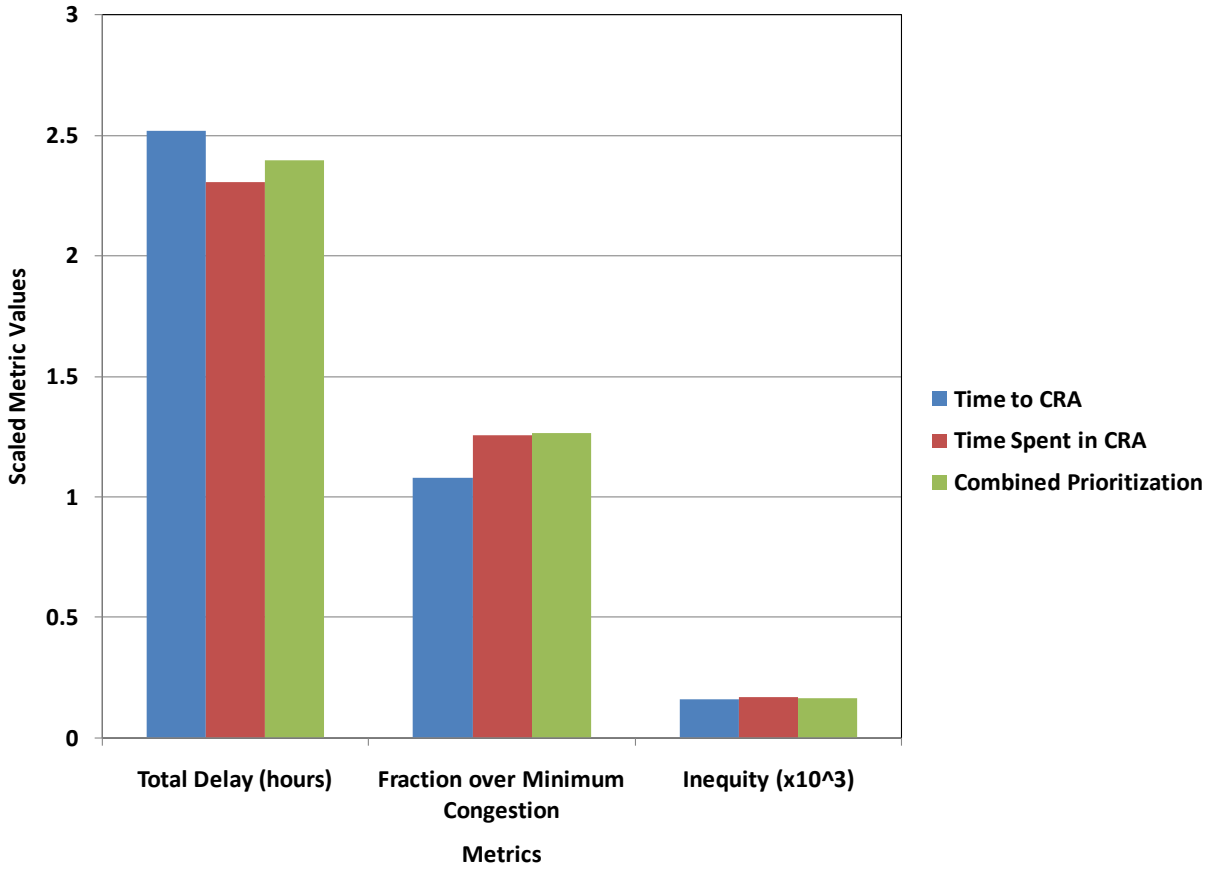

Figure 6-10. Prioritization Criteria Comparison for Minimum Inequity

The ―combined prioritization" achieved a compromise between ―time to CRA" and ―time spent in CRA" for the single metric objective results. In two cases it produces the lowest metric value and in one case the highest metric value and in all other cases, the median metric value. Therefore it provides a more reliable prioritization criterion overall; however it does not necessarily improve the solution quality over the original prioritization criteria. Further investigations into alternative prioritization criteria would be necessary to identify a prioritization criterion and score function pair that consistently produces better solutions for all metrics.

6.3 Multi-Metric Optimization

The previous section shows the impact of the three prioritization criteria when each metric is considered in turn as a single metric objective. In reality, however, achieving balance between multiple objectives is often more important. Therefore, in this section we examine the impact of the prioritization criteria on multi-metric objectives.

6.3.1 Congestion versus Delay

The metrics of delay and congestion are often competing objectives in congestion resolution. In this analysis, we examine how each of the prioritization criteria respond to changes in the relative weightings between these two metrics. The multi-term objective considered is

$$J = k_1 \frac{J_d}{N_d} + k_2 \frac{J_c}{N_c} \qquad \text{Equation 6-7}$$

where $k_1$and k_2 are the weighting factors for delay and congestion, respectively, and $J_d$and J_c are the delay and congestion metrics defined in Section 6.1.3, respectively. To provide meaningful relationships between the weighting factors, the delay and congestion metrics are scaled by constants, N_d and N_c, respectively. Applying the scale factors, the metric values are in units of hours of delay and fraction over minimum congestion.

Table 6-1 shows the various weighting parameters used in the above multi-term objective function to evaluate the impact of varying importance of delay and congestion using the different prioritization criteria.

Table 6-1. Relative Weighting Factor Cases for Multi-Metric Optimization

Case	Weightings
Delay Only	$k_1 = 1, k_2 = 0$
Heavily Emphasized Delay	$k_1 = 10k_2$
Moderately Emphasized Delay	$k_1 = 3k_2$
Equal Delay and Congestion	$k_1 = k_2$
Moderately Emphasized Congestion	$3k_1 = k_2$
Heavily Emphasized Congestion	$10k_1 = k_2$
Congestion Only	$k_1 = 0, k_2 = 1$

Figure 6-11 shows the trend of increasing delay and decreasing congestion shown from left to right, corresponding to the different weightings from delay only to congestion only. For all three prioritization criteria, the ―delay only" solutions provide a small variation in delay (as was shown in Figure 6-8) and a large variation in the resulting congestion. For the ―congestion only" solutions, all prioritization criteria have the same congestion (as is shown in Figure 6-9) and there is only a small variation in delay.

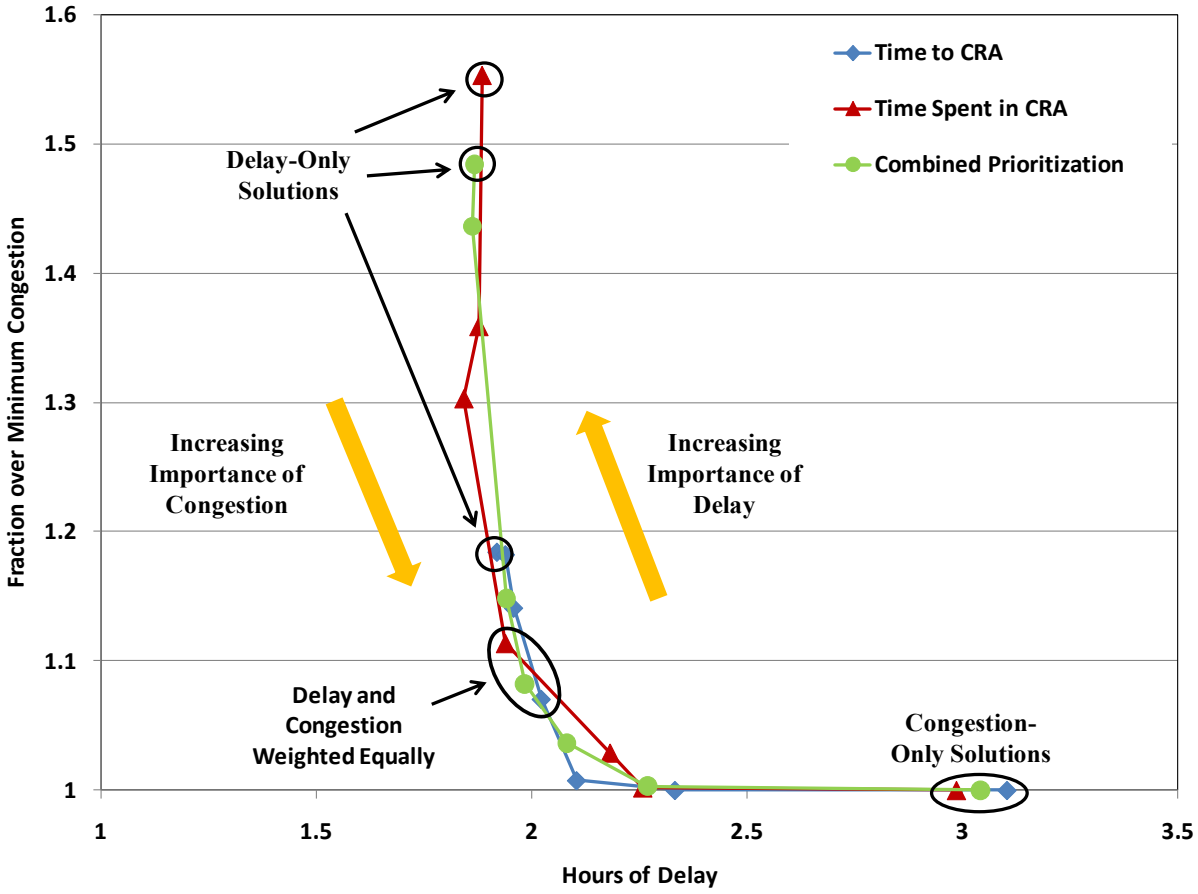

Figure 6-11. Congestion versus Delay

Moving from the ~~"d~~elay only" solutions to the ~~"h~~eavily emphasized delay" solutions shows small increases in delay for large decreases in congestion. Similarly, moving from the ~~"c~~ongestion only" solutions to the ~~"h~~eavily emphasized congestion" solutions shows almost no increase in congestion for significant improvements in delay.

When delay and congestion are equally weighted, the three prioritization criteria provide similar congestion and delay values, which is shown more clearly in Figure 6-12. Examining Figure 6-12 shows that for the ~~"e~~qual delay and congestion" case, ~~"t~~ime spent in CRA" provides the lowest delay and highest congestion. The ~~"t~~ime to CRA" provides the highest delay of the three prioritization criteria, but with the lowest congestion. The ~~"c~~ombined prioritization" provides the balance between the two, as intended. Interestingly, the ~~"c~~ombined prioritization" provides the lowest inequity of all three prioritization criteria, which—depending on priorities—might promote this prioritization over the alternatives.

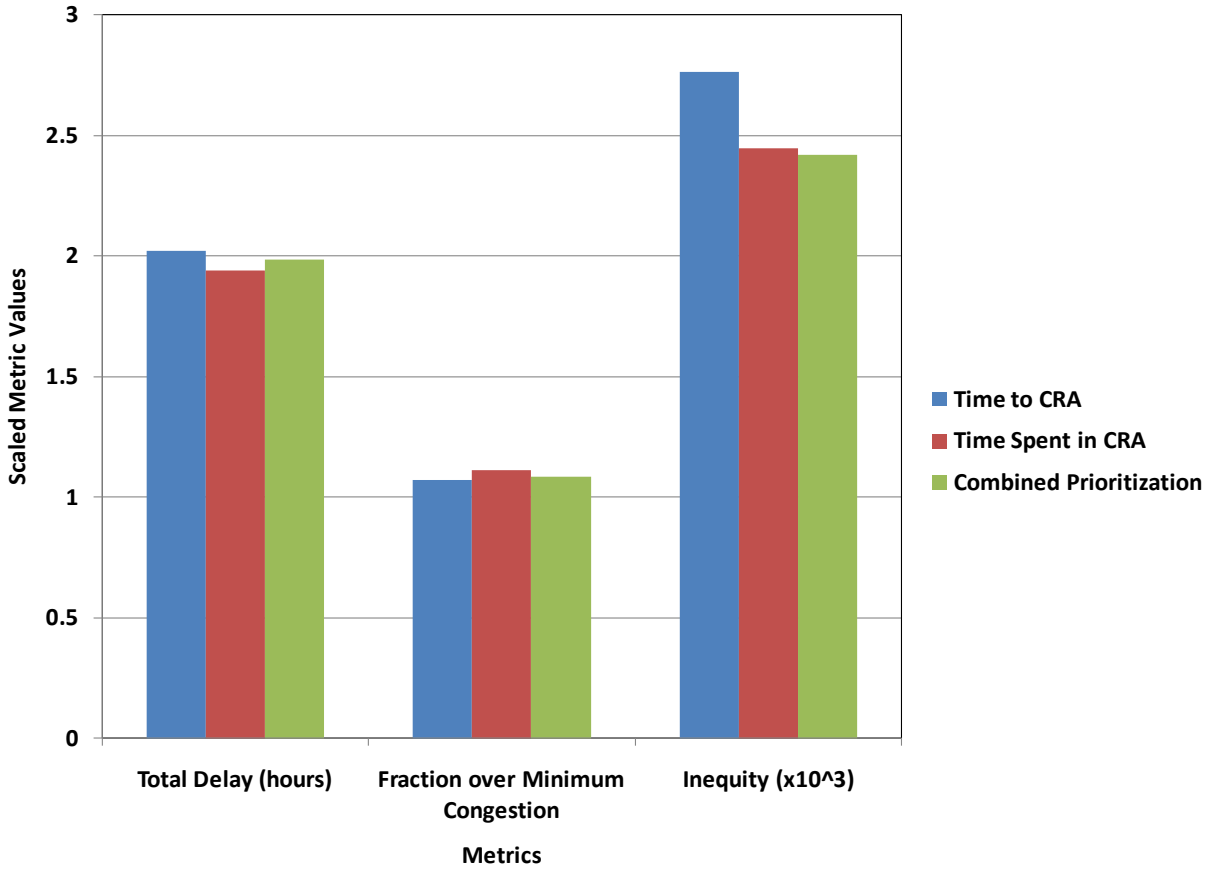

Figure 6-12. Prioritization Criteria Comparison for "Equal Delay and Congestion"

6.3.2 Impact of Inequity

Understanding the trade-offs between delay and congestion is only part of the challenge in congestion management. Additionally, the delays assigned should not unduly impact any given NAS customer. The inequity metric defined in Section 6.1.3 provides one method for measuring the balance of delay distribution. This section investigates the impact of considering inequity directly within the multi-term objective by redefining the objective function as

$$J = k_1 \frac{J_d}{N_d} + k_2 \frac{J_c}{N_c} + k_3 \frac{J_e}{N_e} \qquad \text{Equation 6-8}$$

where the new components of the objective function are k_3, the relative weighting of the inequity component of the objective; J_e, the inequity metric, as defined in Section 6.1.3; and N_e, the scale factor for the inequity metric.

Figure 6-13 shows the minimization of the objective function using the same relationships (and naming conventions) for k_1 and k_2 as described in Table 6-1, and k_3 is set to be 0.1 or a tenth as important as the combination of delay and congestion.

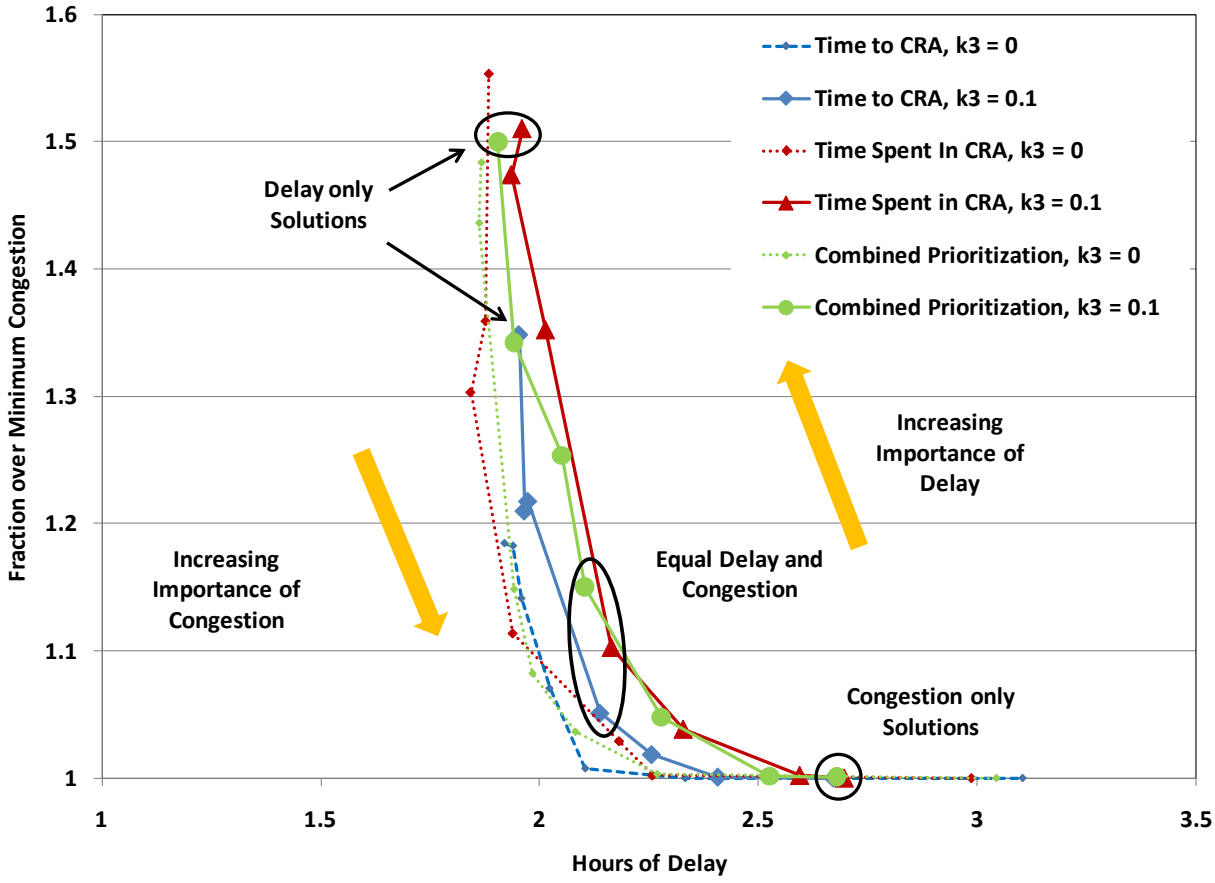

Figure 6-13. Impact of Considering Inequity on Congestion versus Delay

Examining Figure 6-13 reveals that considering inequity, even with a small relative weighting to delay and congestion, changes the overall shape of the delay-congestion trade-off. As before, the same pattern emerges where heavily weighted delay provides a decrease in congestion with only a moderate increase in delay, as compared to the delay-only weighting (with the exception of the ~~–t~~ime spent in CRA," which actually decreases delay as well). Similarly, the heavily weighted congestion does not increase the congestion but does provide a decrease in delay over the congestion-only solutions, albeit not as drastic as in the solutions where inequity is not considered. Finally, when considering the case where delay and congestion are equally weighted, we see a change in the relative trade-offs, where the combined prioritization provides the lowest delay, but the prioritization of time to CRA provides the lowest congestion.

To examine how the prioritization criteria and relative weighting factors in the objective function impact inequity, a more detailed view of three of the weighting factor combinations is presented. For the ―delay only" with inequity case, Figure 6-14 reveals that the lowest delay is produced by the ―combined prioritization" and the lowest congestion is produced by the ―time to CRA." The ―time spent in CRA" produces a slightly higher delay and congestion than the other two prioritization criteria, but significantly reduces the inequity. Decreases in inequity can be achieved through increases in delay as the additional delay can be evenly distributed to reduce inequity; however, the large decrease in inequity is disproportionate to the small increase in delay here, which suggests that the set of flights receiving delays was more equitable.

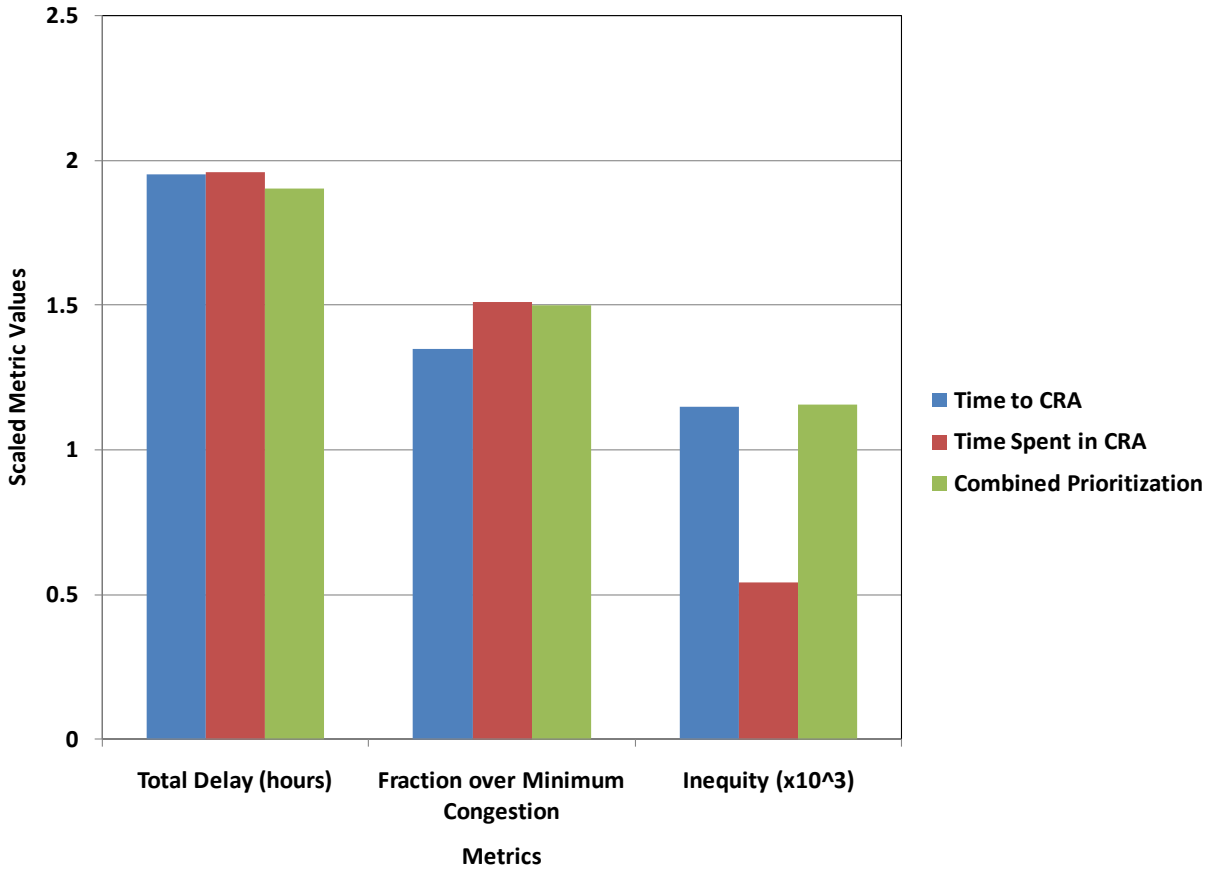

Figure 6-14. Prioritization Criteria Comparison for "Delay Only" with Inequity

For the ―congestion only" with inequity case, Figure 6-15 reveals that the ―time to CRA" produces the minimum congestion, the lowest delay, and the lowest inequity. However, unlike the ―delay only" with inequity case shown in Figure 6-14, the differences in metric values among the three prioritization criteria are very small for both delay and inequity.

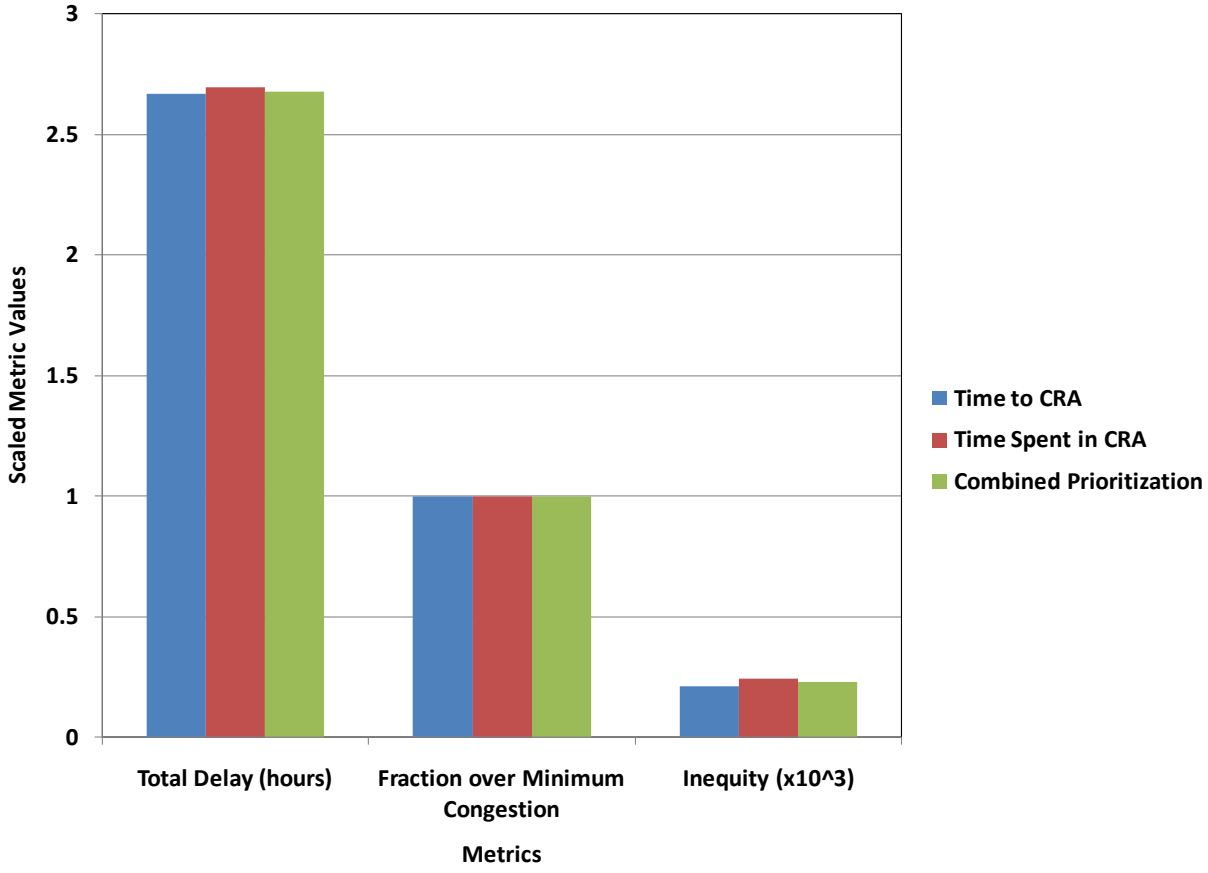

Figure 6-15. Prioritization Criteria Comparison for "Congestion Only" with Inequity

When congestion and delay are equally weighted and inequity is included, Figure 6-16 shows that the ~~"c~~ombined prioritization" provides the lowest delay and the lowest inequity, although the highest congestion. Here, instead of increasing the delay to decrease inequity, the congestion is increased, which describes a situation where a reduction in delay, and therefore a reduction in overall inequity, is achieved at the expense of more closely matching congestion targets.

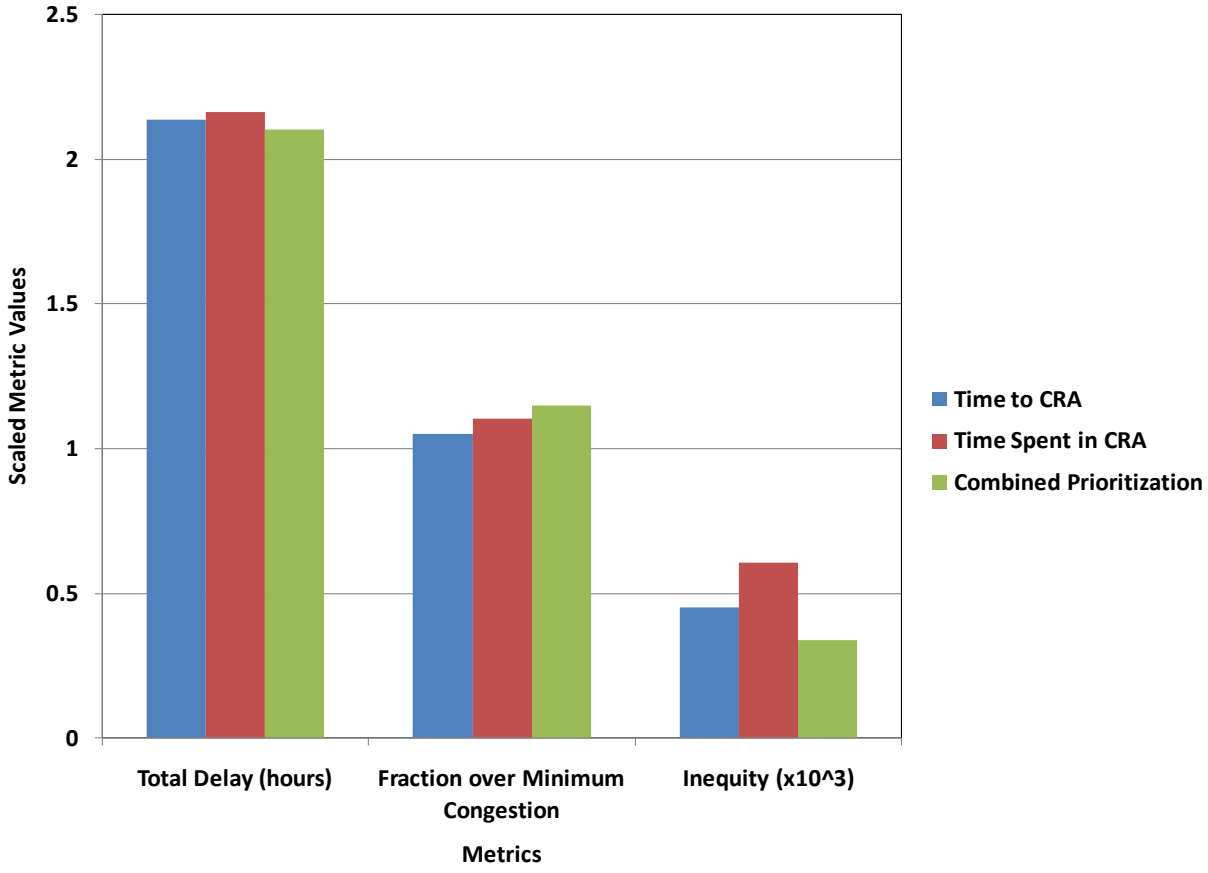

Figure 6-16. Prioritization Criteria Comparison for "Equal Delay and Congestion" with Inequity

In order to more clearly visualize the impact of inequity on metric values, the percent change in metric values from inequity not considered to inequity considered in the objective function are computed for each prioritization criteria and each weighting factor combination. Figure 6-17 shows the percent change in metric values for ~~"~~time to CRA." A negative change shows improvement in the metric value when inequity is considered. Moving from left to right, Figure 6-17 also shows how these changes evolve as congestion is more heavily weighted. Examining Figure 6-17 shows that considering inequity in the objective function always produces significant decreases in the inequity metric and that these reductions become larger as the emphasis on congestion is increased. The reduction of inequity is also paired with an increase in delay for every weighting factor except the ~~"~~congestion only" case. The impact of inequity on congestion varies as congestion is more heavily weighted, in some cases causing an increase, in others a decrease. What is important to note is the magnitude of the changes in the metrics. Considering inequity in the objective decreases inequity between 34% and 95%, while providing maximum increases in delay and congestion of less than 8% and 5%, respectively.

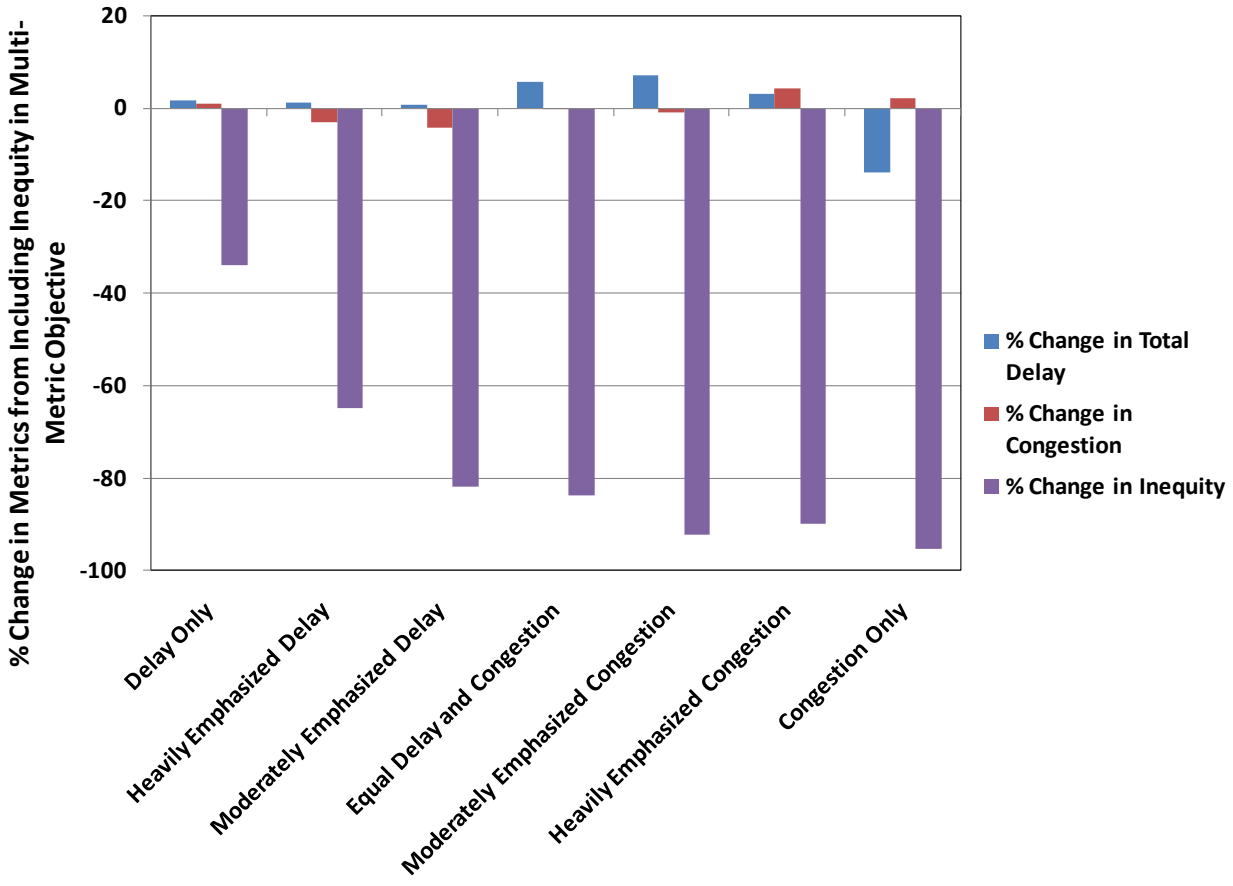

Figure 6-17. Impact of Including Inequity for "Time to CRA"

Figure 6-18 shows the impact of inequity on metric values for the ~~"~~time spent in CRA." Again, we see a significant decrease in inequity; however the trend in decrease varies as congestion is emphasized. An increase in delay results from the consideration of inequity in every case, except for the ~~"~~congestion only" case. And again, the impact on congestion varies. The decrease in inequity ranges between 37% and 93%, while the maximum increases in delay and congestion are less than 15% and 9%, respectively.

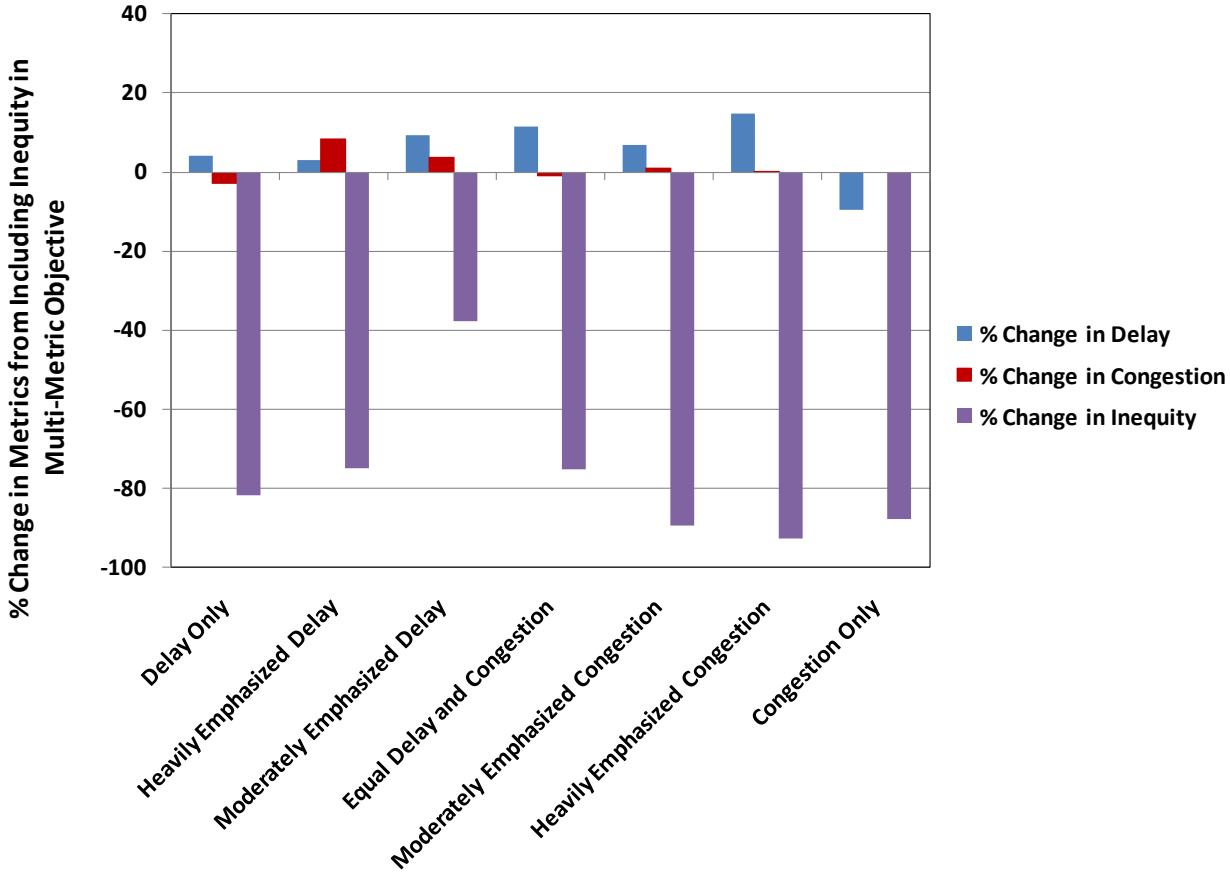

Figure 6-18. Impact of Including Inequity for "Time Spent in CRA"

Figure 6-19 shows the impact of inequity on metric values for the ―combined prioritization." Again, significant decreases in inequity are shown, and the pattern of greater decreases in inequity as congestion is more heavily weighted emerges, like those shown in Figure 6-17 for ―time to CRA." Moving from left to right we see an increase in the change in delay, with the exception of the ―congestion only" case. The impact on congestion from inequity varies. The decrease in inequity ranges between 41% and 93%, while the maximum increases in delay and congestion are less than 12% and 10%, respectively.

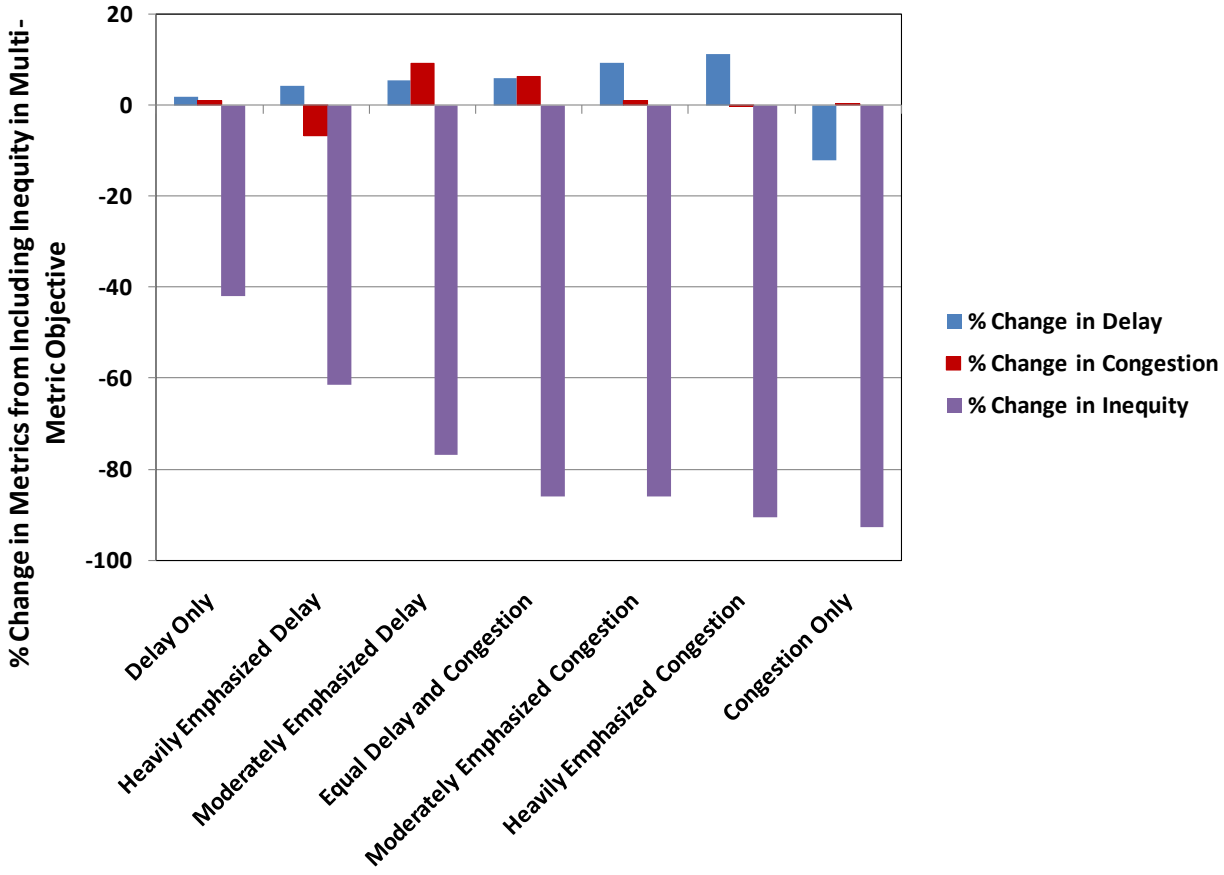

Figure 6-19. Impact of Including Inequity for "Combined Prioritization"

6.4 Summary and Next Steps

In this research, the implementation of GRASP, a stochastic heuristic search procedure, was investigated for its merits in generating improved solutions for airspace congestion problems. Previous research [Wanke et al., 2008b] showed improved solution quality, as compared to results provided by a deterministic greedy heuristic when optimizing single metric objectives. This was extended in FY09 by developing alternative prioritization criteria that provided reliable criteria for producing good solutions. Furthermore, the new and old criteria were used to

generate solutions under a multi-metric objective, to analyze both the quality of the prioritization criteria and trade-offs in the relative priorities of different metrics as they impact airspace congestion resolution performance.

The choice of prioritization criteria, in combination with the selected score functions, defined the relative probability of a flight being selected. These are used to generate a set of possible new priority orders, one of which will provide the best overall solution. By defining the new set of score functions which used the combined prioritization criteria, new distributions were identified that provided a compromise between the individual criteria. The ―combined prioritization" almost always provided the median metric value of the three prioritization criteria. Unlike ―time to CRA" and ―time spent in CRA," which provide the best or worst metric values depending on the specific case, the ―combined prioritization" could be consistently employed to obtain reasonably good solutions regardless of objective function choice. However, there were cases where the combined prioritization criteria did not perform as well. This suggests further research is needed to identify an alternate criterion or another factor to be included in the combined prioritization criteria that will improve solution quality in all cases identified.

This research considered the delay-congestion trade-off in depth, examining multiple relative weightings of delay and congestion using the prioritization criteria selected. The results show that considering heavily emphasized objectives provided almost the same performance of the prioritized metric and significantly improved results for the other metric. This point was further emphasized in the results where inequity was considered. Although, considering inequity in the objective sometimes yielded an increase in delay and congestion for the given weighting, a significant reduction in inequity was obtained. Strikingly, the consideration of inequity in the ―congestion only" case produced a reduction even in delay, further illustrating both the inherent connection of these metrics and the need to evaluate solutions using a multi-metric objective. Further research analyzing alternative weightings of inequity may reveal that a lower weighting on inequity can provide almost identical congestion and delay results while still providing a reduction in inequity.

The analysis and conclusions presented in this research are based on a sample problem that is small in size by design to permit an in-depth analysis of the problem. However, further research into the impacts of GRASP on solution quality using larger, more complex problems is desirable. As the design space increases, the differences in solution quality from different prioritization criteria could become much larger, providing a more clear selection in the choice of heuristic and relative metric weightings.

In addition, it would be enlightening to apply the GRASP method within the simulation described in Section 5, in order to estimate the benefits of using the GRASP method within the overall framework for continual, probabilistic congestion management.

7 Summary

We have described a range of research aimed at developing effective tactical congestion management capabilities. Some of this work is focused on the foundational components of traffic prediction, capacity prediction, and route option generation. The rest is focused on putting the pieces together to develop a practical probabilistic congestion management approach. We made significant progress in all these areas in FY09.

In the demand prediction area, we finished developing the demand model known as ADM2 and extrapolating its demand distributions. The model now provides prediction error distributions for all demand levels and NAS sectors, and improves upon the existing deterministic demand prediction methods.

Our studies of weather-impacted flow capacity greatly improved our understanding of how a set of proposed weather impact metrics work. Enough insight has been gained to propose an initial weather-impacted sector capacity method for operational applications, though the details have yet to be worked out. The initial method would apply a simple 2-dimensional weather coverage metric for low altitude sectors, and a more detailed flow-pattern-based metric for high and super-high altitude sectors.

A new technique for generating reroute options under severe weather conditions was developed and integrated into a decision support prototype. It generates a range of ad-hoc solutions for cases where pre-coordinated or historical routes are not effective or efficient in avoiding weather. This is a key enabler for automated, flight-specific congestion management.

The sequential probabilistic congestion management method developed in FY08 has been converted to a continual method for real-time implementation. Simulation results suggest that a variety of risk management strategies can be used effectively, and can be tuned to achieve a specific balance between congestion risk and incurred cost. The computational requirements are modest and well within reach of today's computers. Also, the deferred congestion resolution maneuver developer proposed in FY08 was simulated, and found to perform better than the simple priority-based congestion resolver used previously.

In the area of developing congestion resolutions, a stochastic partial-optimization method (GRASP) was explored further and found to be effective in developing resolutions that balance congestion resolution effectiveness, incurred delay, and equity of impact among NAS customers. This should improve the performance of the continual probabilistic congestion management approach, at the cost of some added computational complexity.

We hope that the progress reported here brings us closer to deployment of effective tactical congestion management in the NAS.

List of References

Bateman, H., G. Jacobs, L. Rhodes, T. Stewart, S. Zobell, 2009, *Enroute Flow Planning Tool (EFPT) Operational Concept, Evaluation Results, and Next Steps*, MP090203, The MITRE Corporation, McLean, VA.

Davis, E., B. Hargroves, T. Holden, 2005, *The Effect of Convective Weather on Sector Capacity*, MP05W0000200, The MITRE Corporation, McLean, VA.

DeLaura, R., and J. Evans, 2006, *Project Report NASA/A-6: An Exploratory Study of Modeling Enroute Pilot Convective Storm Flight Deviation Behavior*, <http://www.ll.mit.edu/mission/aviation/publications/publication-files/NASA%20A-6.pdf> 11 May 2006 (accessed 21 July 2009), MIT/LL, Lexington, MA.

DeLaura, R., M. Robinson, M. Pawlak, J. Evans, 2008, "Modeling Convective Weather Avoidance In Enroute Airspace," *13th Conference on Aviation, Range and Aerospace Meteorology, 20-24 January 2008, New Orleans, LA* <http://ams.confex.com/ams/pdfpapers/132903.pdf> (accessed 21 July 2009), American Meteorological Society, Boston, MA.

FAA, 2006, *Air Traffic Control System Command Center Severe Weather and Route Management* <http://www.fly.faa.gov/Operations/Strategic_Planning/svrwx_handbook.html> 2006 (accessed 21 July 2009), Federal Aviation Administration (FAA), Washington, DC.

FAA, 2008, *NextGen Implementation Plan* <http://www.faa.gov/about/office_org/headquarters_offices/ato/publications/nextgenplan/0608/> June 2008 (accessed 21 July 2009), FAA, Washington, DC.

Feo, T., and M. Resende, 1995, "Greedy Randomized Adaptive Search Procedure," *Journal of Global Optimization*, Vol. 6, Num 2 <http://www.springerlink.com/content/m57h74q0805g1l43/fulltext.pdf>, March 1995 (accessed 20 July 2009), Springer Netherlands.

GRA, Inc., 2004, *Economic Values for FAA Investments and Regulatory Decisions, A Guide* <http://www.faa.gov/regulations_policies/policy_guidance/benefit_cost/media/050404%20Critical%20Values%20Dec%2031%20Report%2007Jan05.pdf> 31 Dec 2004 (accessed 22 July 2009), prepared for FAA Office of Aviation Policy and Plans.

Hoffman, R., J. Krozel, G. Davidson, D. Kierstead, 2007, "Probabilistic Scenario-Based Event Planning for Traffic Flow Management," *AIAA Guidance, Navigation, and Control Conference and Exhibit, 20-23 August 2007, Hilton Head, SC*, American Institute of Aeronautics and Astronautics (AIAA), Reston, VA.

JPDO, 2007, *Concept of Operations for the Next Generation Air Transportation System*, Version 2.0 <http://www.jpdo.gov/library/NextGen_v2.0.pdf> 13 June 2007 (accessed 21 July 2009), Joint Planning and Development Office (JPDO), Washington, DC.

Krozel, J., J. Mitchell, V. Polishchuk, J. Prete, 2007, "Capacity Estimation for Airspaces with Convective Weather Constraints," *AIAA Guidance, Navigation, and Control Conference and Exhibit, 20-23 August 2007, Hilton Head, SC*, AIAA, Reston, VA.

Martin, B., 2007, "Model Estimates of Traffic Reduction in Storm Impacted En Route Airspace," *7th AIAA Aviation, Integration, and Operations Conference, 18 - 20 September 2007, Belfast, Northern Ireland*, AIAA, Reston, VA.

Masalonis, A., H. Bateman, R. DeLaura, L. Song, N. Taber, C. Wanke, 2008, "Integrated Departure Route Planning," *27th Digital Avionics System Conference, 26-30 October 2008, St. Paul, MN*, Institute of Electrical and Electronics Engineers (IEEE), Los Alamitos, CA.

Mulgund, S., C. Wanke, D. Greenbaum, N. Sood, 2006, "A Genetic Algorithm Approach to Probabilistic Airspace Congestion Management," *AIAA Guidance, Navigation, and Control Conference and Exhibit, 21-24 August 2006, Keystone, CO*, AIAA, Reston, VA.

Nilim, A., L. El Ghaoui, V. Duong, 2003, "Multi-Aircraft Routing and Traffic Flow Management Under Uncertainty," *5th USA/Europe Air Traffic Management R&D Seminar, 23-27 June 2003, Budapest, Hungary* <http://www.atmseminar.org/past-seminars/5th-seminar-budapest-hungary-june-2003/papers/paper_101/> (accessed 22 July 2009).

Song, L., C. Wanke, D. Greenbaum, 2007a, "Predicting Sector Capacity for TFM," *7th USA/Europe Air Traffic Management Research and Development Seminar, 2-5 July 2007, Barcelona, Spain* <http://www.atmseminar.org/past-seminars/7th-seminar-barcelona-spain-july-2007/papers/paper_038/> (accessed 21 July 2009).

Song, L., C. Wanke, D. Greenbaum, D. Callner, 2007b, "Predicting Sector Capacity under Severe Weather Impact for Traffic Flow Management," 2007-7887, *7th AIAA Aviation Technology, Integration, and Operations Conference, 18-20 September 2007, Belfast, Northern Ireland*, AIAA, Reston, VA.

Sood, N., S. Mulgund, C. Wanke, D. Greenbaum, 2007, "A Multi-Objective Genetic Algorithm for Solving Airspace Congestion Problems," *AIAA Guidance, Navigation, and Control Conference and Exhibit, 20-23 August 2007, Hilton Head, SC*, AIAA, Reston, VA.

Steiner, M., R. Bateman, D. Megenhardt, J. Pinto, 2009, "Evaluation of Ensemble-Based Probabilistic Weather Information for Air Traffic Management," *Aviation, Range and Aerospace Meteorology (ARAM) Special Symposium on Weather – Air Traffic Management Integration, 11-15 January 2009, Phoenix, AZ* <http://ams.confex.com/ams/pdfpapers/143781.pdf> (accessed 5 August 2009), American Meteorological Society, Boston, MA.

Taylor, C., M. Song, D. Klabjan, O. de Weck, D. Simchi-Levi, 2007, "A Mathematical Model for Interplanetary Logistics," *Logistics Spectrum,* Vol. 41, Num. 1, January 2007, The International Society of Logistics, Hyattsville, MD, pp. 23-33.

Taylor, C., and C. Wanke, 2008, "A Generalized Random Adaptive Search Procedure for Solving Airspace Congestion Problems," *AIAA Guidance, Navigation, and Control, 17-22 August 2008, Honolulu, HI*, AIAA, Reston, VA.

Volpe, 2002, *Enhanced Traffic Management System (ETMS) Reference Manual, Version 7.5*, Report No. VNTSC-DTS56-TMS-004, Volpe National Transportation Systems Center, U.S. Department of Transportation, Cambridge, MA.

Wanke, C., and D. Greenbaum, 2008, "Sequential Congestion Management with Weather Forecast Uncertainty," *AIAA Guidance, Navigation, and Control Conference and Exhibit, 18-21 August 2008, Honolulu, HI* <http://www.mitre.org/work/tech_papers/tech_papers_08/08_1174/08_1174.pdf> (accessed 6 August 2009), AIAA, Reston, VA.

Wanke, C., N. Taber, S. Miller, C. Ball, and L. Fellman, 2003, "Human-in-the-Loop Evaluation of a Multi-Strategy Traffic Management Decision Support Capability," 5th *USA/Europe Air Traffic Management R&D Seminar*, Budapest, Hungary, June 2003.

Wanke, C., L. Song, S. Zobell, D. Greenbaum, S. Mulgund, 2005a, "Probabilistic Congestion Management," *6th USA/Europe Seminar on Air Traffic Management R&D, 27-30 June 2005, Baltimore, MD* <http://www.atmseminar.org/past-seminars/6th-seminar-baltimore-md-usa-june-2005/papers/paper_008/> (accessed 29 July 2009).

Wanke, C, S. Zobell, L. Song, 2005b, "Probabilistic Airspace Congestion Management," *5th Aviation Technology, Integration, and Operations Conference, 26-28 September 2005, Arlington, VA*, AIAA, Reston, VA.

Wanke, C., N. Taber, D. Greenbaum, C. Jackson, G. Jacobs, A. Masalonis, L. Rhodes, L. Song, T. Stewart, C. Taylor, S. Zobell, 2008a, *Research Results for Advanced Congestion Prediction and Resolution Capabilities,* MP080087, The MITRE Corporation, McLean, VA.

Wanke, C., N. Taber, D. Greenbaum, C. Jackson, G. Jacobs, A. Masalonis, L. Rhodes, L. Song, C. Taylor, S. Zobell, 2008b, *Research Plan for Advanced Congestion Prediction and Resolution Capabilities: Fiscal Year 2009 and Beyond,* MP080088, The MITRE Corporation, McLean, VA.

Wanke, C., 2009, "Continual, Probabilistic Airspace Congestion Management," *AIAA Guidance, Navigation, and Control Conference, 10-13 August 2009 Chicago, IL*, AIAA, Reston, VA.

Weber, M., J. Evans, M. Wolfson, R. DeLaura, B. Moser, B. Martin, J. Welch, J. Andrews, D. Bertsimas, 2006, "Improving Air Traffic Management During Thunderstorms," *12th Conference on Aviation, Range, and Aerospace Meteorology, 29 January - 2 February 2006, Atlanta, GA* <http://www.ll.mit.edu/mission/aviation/publications/publication-files/WW-12668.pdf> (accessed 12 July 2009), American Meteorological Society, Boston, MA.

Zobell, S., D. Greenbaum, L. Song, D. Tiffany, C. Wanke, 2005, *Algorithms and Capabilities for Automated Probabilistic Congestion Management,* MTR05W037, The MITRE Corporation, McLean, VA.

Glossary

2-D	two-dimensional
3-D	three-dimensional
4-D	four-dimensional
ACES	Adaptation Controlled Environment System
ADM	Aggregate Demand Model
AIAA	American Institute of Aeronautics and Astronautics
ARAM	Aviation, Range and Aerospace Meteorology
ARTCC	Air Route Traffic Control Center
ATL	Atlanta Hartsfield International Airport
ATM	Air Traffic Management
CAASD	The MITRE Corporation's Center for Advanced Aviation System Development
CIWS	Corridor Integrated Weather System
CMA	Congestion Management Area
CRA	Congestion Resolution Area
CRCT	Collaborative Routing Coordination Tools
CWAM	Convective Weather Avoidance Model
DOC	direct operations cost
EDT	Eastern Daylight Time
EFPT	En route Flow Planning Tool
ETMS	Enhanced Traffic Management System
FAA	Federal Aviation Administration
FOG	Flight Option Generation
FY	Fiscal Year
GRASP	Generalized Random Adaptive Search Procedure
IDRP	Integrated Departure Route Planning
IEEE	Institute of Electrical and Electronics Engineers
JPDO	Joint Planning and Development Office

km	kilometer
LAT	look-ahead time
MAP	Monitor/Alert Parameter
MIT/LL	Massachusetts Institute of Technology Lincoln Laboratory
MOIE	Mission Oriented Investigation and Experimentation
MSR	MITRE-Sponsored Research
MTBD	mean time before departure
NAS	National Airspace System
NCAR	National Center for Atmospheric Research
NextGen	Next Generation Air Transportation System
PDF	probability distribution function
PMF	probability mass function
SD	standard deviation
TFM	Traffic Flow Management
TFMS	Traffic Flow Management System
UTC	Coordinated Universal Time
VIL	Vertically Integrated Liquid
VIL3+	VIL at and above level three
WAAF	weather avoidance altitude field
WP	Work Package
ZDC	Washington ARTCC
ZID	Indianapolis ARTCC
ZNY	New York ARTCC
ZOB	Cleveland ARTCC