Assessing NAS Performance: Normalizing for the Effects of Weather

Michael B. Callaham, Sc.D. * (mcallaha@mitre.org), James S. DeArmon* (jdearmon@mitre.org), Arlene M. Cooper* (cooper@mitre.org), Jason H. Goodfriend, Ph.D.+ (jason.goodfriend@lmco.com), Debra Moch-Mooney* (dmooney@mitre.org), George H. Solomos* (gsolomos@mitre.org)

*The MITRE Corporation, 7515 Colshire Drive, McLean, VA, 22102-7508, USA

+The MYTA Corporation, 11400 Rockville Pike, Suite 210, Rockville, MD 20852, USA

1 Abstract

Frequently, it is useful to compare a National Airspace System (NAS) performance measure over different time intervals. Such a comparison can be used to determine the effect of NAS improvements. The problem with conducting this comparison is that different time intervals may experience very different weather, which can have a large impact on NAS performance. Without taking into account the differences in weather, this comparison is tantamount to comparing "apples and oranges."

This paper addresses this problem of weather normalization in two ways. The first method is to define a Weather Impacted Traffic Index (WITI) that assigns a weighted scalar value to a given day. An en route WITI normalizes for the potential impact of severe convective weather on en route performance. An en route/terminal WITI consists of the en route WITI plus terminal weather variables. Thus, days in different time intervals that had similar weather impacts can be grouped together according to their WITI values and then fairly compared. The en route/terminal WITI is utilized to compare NAS performance of Spring/Summer 1999 with Spring/Summer 2000.

The second method of weather normalization employs cluster analysis to determine the discrete weather *day-types* with respect to severe en route weather. This approach includes the determination of traffic *day-types*. This classification scheme differs from the WITI, which is a continuous-scaled index. Initial trials at model building using these weather-traffic day-type classifications provide encouraging results. Correlation coefficients on the order of 65 percent indicate that a non-trivial amount of variability in a performance measure can be attributed to weather and traffic.

2 Introduction

A recurring challenge in National Airspace System performance assessment is the problem of accounting for the effect of weather on air traffic. If two days have different weather conditions, the weather effects must be partitioned before a fair comparison of other effects, e.g., traffic level, air traffic management performance, automation enhancements, can be undertaken.

There are two other potential applications of techniques for normalizing performance for effects of weather. One is turning a forecast of tomorrow's weather into a forecast of tomorrow's NAS performance. Another is helping the Severe Weather Desk at the Air Traffic Control System Command Center (ATCSCC) make strategic decisions. For example, if the ATCSCC knows what today's performance is expected to be, given the weather, they can evaluate their plans for handling NAS traffic based on that expectation.

This study uses two different approaches for normalizing the weather's impacts on NAS performance: the Weather Impacted Traffic Index approach and the weather-traffic day-type clustering approach.

2.1 The Weather Impacted Traffic Index Approach

The first method for weather normalization is to define a Weather Impacted Traffic Index (WITI) that assigns a weather value to a given day. Specifically, the WITI normalizes for the potential impact of severe convective and terminal area weather on performance. Thus, days in different time intervals can be grouped together according to their WITI values and then compared.

The en route WITIs are calculated by overlaying a grid on the United States (U.S.) and assigning a weight to each grid cell based on the potential

-

impact on NAS performance if that cell were to experience severe convective weather. Since severe convective weather has the greatest effect on en route air traffic during the spring and summer months, the scope of this analysis is during these months.

Terminal weather variables are included in the model to account for the impact of airport weather on NAS performance. Various terminal weather variables were analyzed, but only those that showed a significant relationship with NAS performance are included in the model.

Section 3 explains the approach used to construct the en route and terminal area WITIs. It describes the data used in the analysis, including the NAS performance measures. The results of the analysis are then presented. The utility of the analysis is demonstrated by comparing NAS performance for Spring/Summer 1999 with Spring/Summer 2000.

2.2 The Clustering Approach

The second method analyzed for weather normalization uses cluster analysis to determine the contribution of weather to NAS performance. This method also uses cluster analysis on scheduled traffic to determine traffic day-types. Cluster analysis provides a classification scheme that determines groups or day-types with respect to both traffic and weather. This classification scheme differs from the WITI, which is a continuous-scaled index.

The two day-type schemas (traffic classification and weather classification) were used in regression modeling. The basic assertion of the modeling effort is that, at a high level, system performance is a function of traffic, weather and other factors, e.g., human decisions, Traffic Flow Management (TFM) strategy, and responses to irregular operations.

NAS Impact = f (Traffic Level and Patterns, Weather, Other Factors)

A response variable (NAS performance measure) is required to evaluate the NAS impact in the relationship noted above. Through discussions with airline operations staff, it was learned that flight cancellations, diversions, and long delays are all indicative of serious schedule disruption. Hence, a composite measure was created, a *flight impact value* (FIV), which is a linear combination of the proportions of daily flights subject to the various ill effects. The FIV is the single performance measure used in the classification and clustering approach.

As a trial of the proposed modeling, a specific major U.S. carrier is considered. This carrier is designated herein as *ABC*.

Section 4 develops these ideas further and provides more detail on the approach and associated results.

3 The WITI Approach

3.1 En Route WITIs

The ability to utilize information about convective weather that could impact the en route environment has increased due to the recent availability of the National Convective Weather Forecast (NCWF) database. In addition to providing forecasts, this database provides historical information on the actual location of severe convective weather. In order for convective weather to qualify as severe, it must register a level of 3 or above on a scale of 6 levels defined for Next Generation Weather Radar (NEXRAD) data. The location of the severe weather is provided in the form of polygons with specified latitudes and longitudes. No altitude information is given—severe weather typically is found between altitudes of approximately 18,000 to 50,000 feet above mean sea level (MSL)—altitudes used by commercial and military air traffic. These polygons provide the basis for the development of the WITIs.

The first step in the WITI development is to overlay a grid across the United States. The basic concept is to assign a weight to each cell in the grid, such that this weight is a measure of the potential impact on air traffic performance should severe convective weather occur in this cell.

To assign the weights, a *good weather day* was selected (in terms of the absence of major storms and cloud cover). The weight for each cell was initially based on the volume of air traffic in that cell. The traffic volume was estimated to be the number of Track Update (TZ) (position report) Enhanced Traffic Management System (ETMS) messages for that cell. Since traffic counts are dependent on the time of day, 24 grids, one for each hour of the Greenwich Mean Time (GMT) day, were developed, and the weight became the appropriate number of TZ messages for each cell for each hour. In order to focus on en route traffic, TZ messages were counted only for aircraft positions that were above 18,000 feet

MSL. Also, in order to reduce redundancy, a TZ message was ignored if it occurred within 5 minutes of the previous message, assuming that both TZ messages were for the same aircraft.

The reasoning behind using this cell-weighting scheme is that the number of TZ messages roughly captures the number of flights that would have flown through the cell in the absence of severe convective weather. That is, the more flights that want to fly through the cell, the more potential disruption there is to performance in the presence of severe convective weather. Figure 1 displays the grid for one hour of a good weather day. The cells with relatively higher numbers represent those locations with greater air traffic density. The cell size for the grid in the figure is approximately 0.5 degree longitude by 0.5 degree latitude.

The NCWF data are reported every 5 minutes. That is, every 5 minutes there is supposed to be a new NCWF file—called a *report* in this paper—that lists the polygons outlining severe convective weather across the United States, although there are time periods with no reports. There can be at most 288 reports for a given day (12 reports per hour times 24 hours per day).

Several WITIs were tested and based on the results, the analysis focused on two WITIs: WITI A and WITI B. The difference between WITI A and WITI B is that WITI B places more emphasis on blockage of east-west traffic due to the weather. The more severe weather in a particular column of the grid and the fewer gaps in the weather, the higher is the score associated with that column. An example of the calculation of Score A and Score B is provided based on the sample grid illustrated in Figure 2. For convenience, this figure displays a notional grid for a given report (as opposed to the real grid which has many more cells) and notional weights indicated by the number in each cell. A severe weather cell is indicated in the figure by the presence of rain clouds. Each rain cloud indicates the presence of severe convective weather in the associated cell.

3.1.1 WITI A

The first step to calculate WITI A is to calculate *Score A* for each report. To obtain this score, the grid corresponding to the hour of the report is utilized. Score A is simply the sum of the weights of the grid for those cells that intersect with at least one NCWF polygon (i.e., a cell where severe weather is occurring).

Score A = 8 + 1 + 10 + 10 + 1 +6 + 8 + 10 = 54

The scores for the day are then averaged with respect to the number of reports. Hence, WITI A is the average value per report of Score A for the day.

3.1.2 WITI B

The first step to calculate WITI B is to calculate *Score B* for each report. To obtain this score, the grid corresponding to the hour of the report is utilized. Score B is the sum of the column scores. Each column score is the sum of the traffic-contiguity weights for those cells in the grid column that intersect with at least one NCWF polygon (i.e., cells where severe weather is occurring). The traffic-contiguity weight of a severe-weather cell is its weight (described above) times the number of cells in the sub-column of contiguous severe-weather cells containing the cell. The scores for the day are then averaged with respect to the number of reports. Hence, WITI B is the average value per report of Score B for the day.

Score B = 8x4 + 1x4 + 10x4 + 10x4 + 1x1 +
6x1 + 8x1 + 10x1 = 141

As currently implemented, the position reports do not capture the complexity of air traffic and the workload on air traffic controllers. The impact of severe weather is felt not just in areas with high volume, but also in areas of high complexity (e.g., transition or holding air traffic). Since severe weather in a cell with 30 converging aircraft will have more of an operational impact than weather in a cell with 30 aircraft flying straight and level, future analysis includes applying additional weights for airspace complexity. The NAS Operational Evolution Plan [1] describes en route congestion problems and solutions. The concept of *choke points* is used as a mechanism to define the fast-track priorities for improvements in the efficiency of the aviation system. Choke points can also be used to identify particularly complex airspace (airspace where a little disruption goes a long way). Choke point locations could be identified on the en route WITI grid and traffic counts could then be multiplied by *complexity factors* to reflect the additional challenges associated with re-routing traffic in those areas. The application of complexity factors will be addressed in the next phase of the project.

-

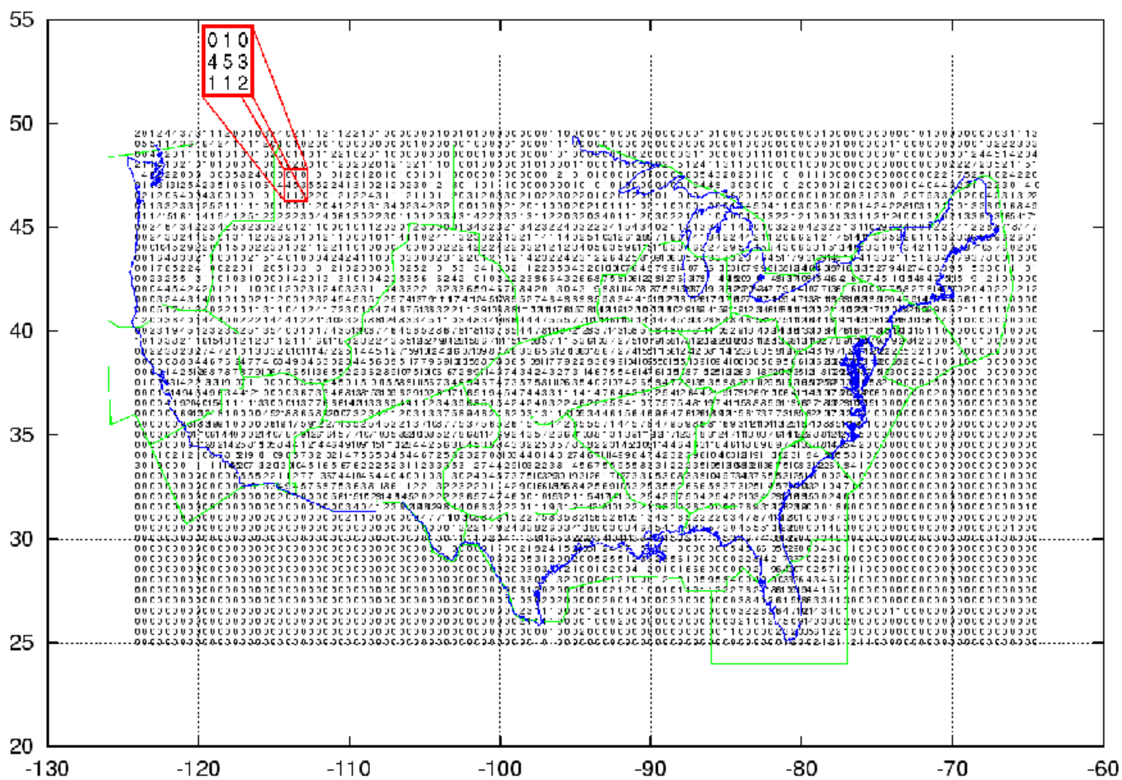

Figure 1. Weighted Grid

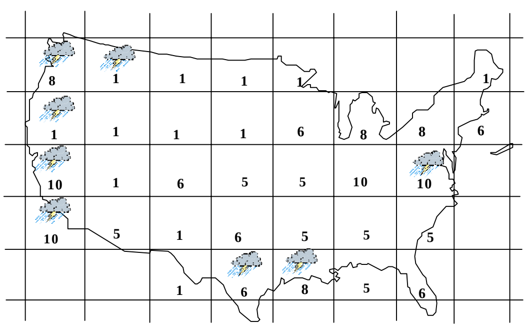

Figure 2. Notional Grid for Demonstrating Calculation of Scores

3.2 Inclusion of Terminal Weather

In order to provide a more complete picture of the weather's impact on NAS performance, terminal area weather variables were also analyzed for inclusion in the model. These variables were analyzed to account for the impact of airport weather conditions on NAS performance. The Surface Weather Observations from the National Climatic Data Center (NCDC) were used to obtain terminal weather data.

Specifically, daily observations of ceiling and visibility for 49 airports—the 50 busiest airports in the NAS in terms of ETMS operations, less Metropolitan Oakland International Airport, for which 1999 NCDC data were not available in electronic form—were utilized to determine the minutes of Instrument Meteorological Conditions (IMC) for the given day at each airport. It was hypothesized that NAS performance is reduced under IMC conditions as compared to Visual Meteorological Conditions (VMC).

The effects of terminal area wind on NAS performance were also considered for this study. The impacts of wind direction (as it relates to airport configuration) and wind speed on airport capacity and delay were analyzed to determine if they should be included in the model. The Aviation System Performance Measurements (ASPM) hourly data were used to conduct a preliminary analysis of performance, wind angle (direction), airport configuration and wind speed to determine which, if any of the wind variables,

-

should be included in the model. The results of the analysis indicate that wind speed does show a significant relationship with NAS performance whereas wind direction does not. Based on this analysis, the daily average wind speed for the 49 airports are included as independent variables in the model in addition to the IMC independent variables.

The inclusion of en route and terminal weather independent variables in the model provides a regression equation of the form:

> **Performance Measure = f (en route WITIs, IMC for each of 49 airports, WIND for each of 49 airports)**
>
> where IMC = daily minutes of IMC,
>
> and WIND = daily average wind speed for the airport

3.3 NAS Performance Measures

Several NAS performance measures were considered for this study. These performance measures include:

- Average arrival delay for flights with more than 15 minutes of arrival delay
- Average block time
- Average departure delay for flights with positive departure delay
- Average departure delay for flights with 15 minutes or more of departure delay
- Operations Network (OPSNET) weather delay rate

Results for average arrival delay greater than 15 minutes are provided in this paper. Average arrival delay for flights with more than 15 minutes of arrival delay is obtained from U.S. Department of Transportation Airline Service Quality Performance (ASQP) data. Arrival delay is defined to be the difference between the actual arrival at the destination gate and the Computer Reservation System (CRS) scheduled arrival time at the destination gate. Therefore the *average arrival delay for flights with more than 15 minutes of arrival delay* is defined to be the average arrival delay, where the average is taken over those flights that had an arrival delay of greater than 15 minutes.

3.4 Analysis Results

Figure 3 illustrates the use of a WITI for comparing performance on different days. Suppose that one wishes to compare average arrival delay greater than 15 minutes on September 9, 1999, with the same measure for other May-September 1999-2000 days. The figure shows, on the vertical axis, values of this delay measure for 271 days during May to September 1999 and May to September 2000. The thick horizontal line shows the mean delay for this sample of days (53.6 minutes), and the thinner horizontal lines show the mean delay plus and minus the standard deviation (10.4 minutes). The delay for September 9, 1999, ranks 13th, at the 96th percentile. It is 21 minutes—more than two standard deviations above average.

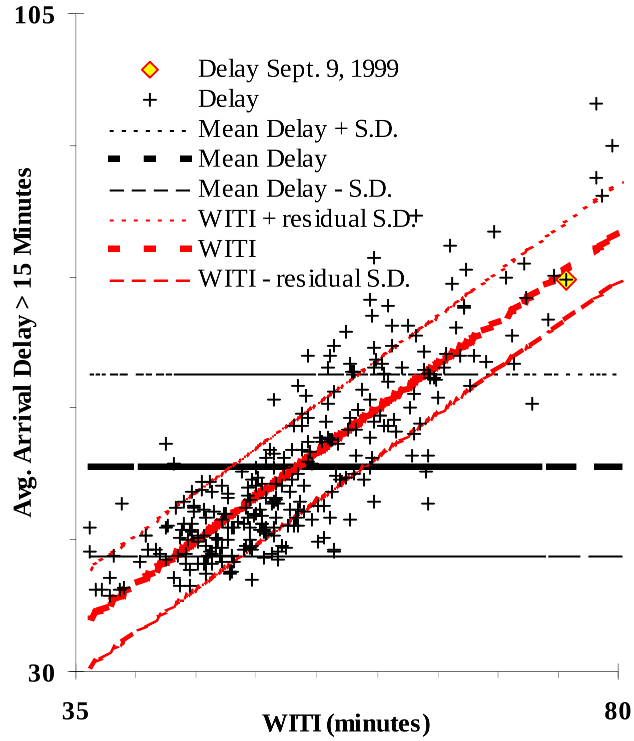

Figure 3. Comparing Average Arrival Delays Greater than 15 Minutes

For each day, the horizontal axis shows the value of a WITI derived for this delay measure. The WITI is the conditional mean delay, given the values of the en route WITIs and weather variables. It has units of delay: minutes. It was derived by regression, as described in section 3.2. The thick diagonal line shows the expected delay given the WITI. The thinner diagonal lines show the conditional mean delay plus and minus the standard deviation of residuals (estimation errors), which is 5.6 minutes. Given the WITI for the day, the delay for September 9, 1999 is 0.2 residual standard deviations below the

-

conditional mean. It ranks 4th among the 7 days that have a WITI differing by at most three minutes from the WITI for September 9. Thus the September 9 delay is unremarkable after normalization for weather using this WITI.

To avoid overfitting the model, the number of explanatory variables was reduced so as to minimize Akaike's Information Criterion (AIC) [2] using the stepwise regression capability of the R software package [3]. Figure 4 shows, for each revision date from 1999-09-30 to 2000-09-30, the R^2 and adjusted R^2 of the full- and reduced-dimensional models. It also shows the number of parameters retained in the reduced-dimensional model.

3.5 Nonlinear Transformation

It is evident from Figure 3 that the average arrival delays greater than 15 minutes are not normally distributed. The residuals (delay less WITI) are also not normally distributed, which precludes the justified use of diagnostic tests such as t-tests for the significance of explanatory variables. However, the residuals of a 101-parameter model fit to the logarithms of delays are normally distributed (p=0.29, Shapiro-Wilk test), and the exponentials of the fitted values of

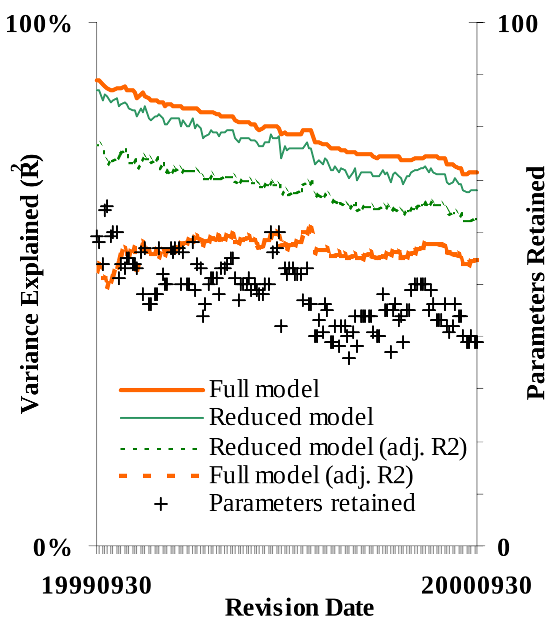

Figure 4. Delay Variance Explained by Full- and Reduced-Dimensional Models

the logarithms explain 72 percent of the delays, a bit more than the R^2 of the linear model (0.71).

3.6 Spring/Summer Comparison

Figure 5 shows the empirical cumulative distribution of weather-adjusted daily average arrival delays greater than 15 minutes in May to September (*Spring/Summer*) 1999 and the corresponding distribution for Spring/Summer 2000. Comparison indicates that the distributions are very similar. The maximum (over all delays) difference between the distributions is 0.065. If one assumes that each day's delay was a random sample from an underlying probability distribution for the year, an exact Smirnov test would not find the distributions to be significantly different (p=0.912). That is, the differences might be due to chance.

The adjustment for weather was performed by subtracting from each daily delay the corresponding WITI—i.e., the expected delay given the day's weather data—calculated using the 101-parameter regression model described above, fit to all Spring/Summer 1999 and 2000 data.

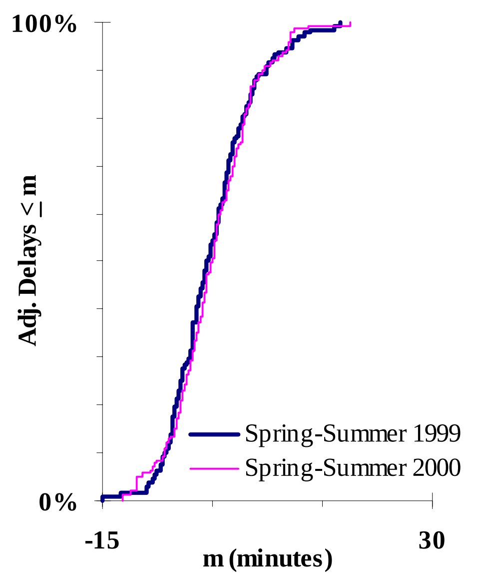

Figure 5. Distributions of Weather-Adjusted Daily Average Arrival Delays Greater than 15 Minutes

The adjusted delay therefore is the component of delay not attributed to weather by the regression model. A negative adjusted delay indicates that delay was lower than average, given the weather.

4 Weather and Traffic Cluster Analysis

In this section, a different approach is taken in order to discern the level of contribution of various factors to NAS performance. Using cluster analysis, both severe weather day-types and traffic day-types are created.

4.1 Creating Severe Weather Day-Types

The severe weather polygons of the NCWF database are used. After removing data anomalies, a set of days from April-October of both 1999 and 2000 were compiled for analysis. This weather data was reduced and formatted to represent days of severe weather in the NAS.[1]

Severe weather polygons are reported in the NCWF data about every 5 minutes. Represented graphically, it would be typical to see a trace of polygons, in 5-minute transitions, slightly *morphing* and moving west-to-east, consistent with the prevailing winds in the Continental United States (CONUS). In order to group similar days with respect to severe weather data, the essentially *non-regular data* needed to be *regularized*. The location, shape, extent, and movement over time of the severe weather areas were needed. To accomplish this task, a gridding structure of 50 x 50 nautical miles (nmi) square cells was overlaid, and information on a per-cell basis was captured. The regularity of the gridding structure allows a consistent comparison across days. (And note the obvious trade-offs regarding fineness of gridding, too coarse a gridding means insufficient representation of the information, whereas too fine a gridding implies excess information representation and undue computer processing time and space.)

It is easy to see how a gridding structure captures information about weather location, shape, and extent, but representing weather movement over time is a bit more involved. For this, the 17-hour daytime period in the CONUS, 6 a.m.—11:00 p.m., was divided into four roughly equal time periods lasting {4, 4, 4, and 5} hours.[2] If the weather were moving west-to-east, it might be captured in the gridding structure as covering successive, adjacent grid cells for a given time period, and then, for the next time period, covering other grid cells as it moved. See Figures 6 and 7 for graphical representations of, respectively, weather polygons, and their gridded representations. Note the color coding for time periods and polygon sequences morphing and moving.[3]

Since this study focuses on severe weather as it impacts commercial air traffic, it is not logical to consider all grid cells as equally important: severe weather over Montana does not impact traffic as it does over Georgia. After some experimentation, a weighting scheme tied closely to intended flight traffic was used. The top 50 airport pairs, with respect to frequency of directed flight legs (ASQP data from 5/1/00), was used. The paths of these directed flight legs (simple direct circle routing) were plotted indicating which grid cells in the gridding structure were intersected. The intersected grid cells were then weighted by the frequency of flight leg, mapped into percentiles of that distribution. For example, New Orleans to Houston has fewer flight legs than does Minneapolis to Chicago, and so the grid cells along the former path are weighted relatively less, at the accuracy of one part in a hundred.

After grid weighting, groupings of the 239 severe weather days were computed. The data were arranged as a matrix, in which rows were days and columns were the severe weather cell attributes, i.e., a linearization of the rectangular gridding structure. In this linear concatenation, weighting of cells is accomplished by making additional copies of higher-weighted cells. Cells with zero weights are discarded in this step. Note also that four representations of this described row data structure are needed for the four time periods of the subject day.

[1] This study assumes that a reasonable entity for study is a daytime epoch in the CONUS, which is defined as 6 a.m. to 11:00 p.m. eastern time. This seemed logical for this exploratory analysis, since the air carriers are oriented to daily operations: fulfilling flight schedules and getting airframes in place for the next day, and, in the face of irregular operations, delivering passengers to destinations at worst by the end of the day, to avoid the cost of overnight lodging for stranded passengers.

[2] Although not derived scientifically, this breakdown seemed to provide a granularity consistent with airline flight planning, route planning is done 1 to 2 hours before departure, and flight times en route are typically 1 or 2 hours long.

[3] To avoid an overabundance of data, only the first observation of each hour (out of a possible total of 12) is used.

This matrix is appropriate input to a *clustering algorithm*. Clustering is a statistical procedure that determines similarity of observations, using

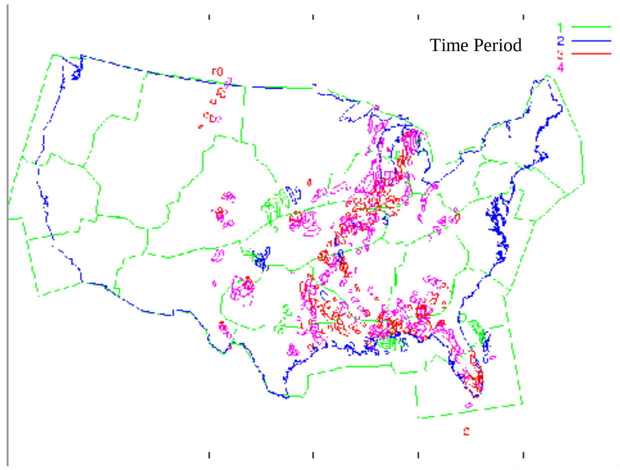

Figure 6. Severe Weather Polygons, 6/11/99
(Shade or color indicates time—1: 6am-10am, 2: 10am-2pm, 3:2pm-6pm, 4: 6pm-11pm Eastern Time)

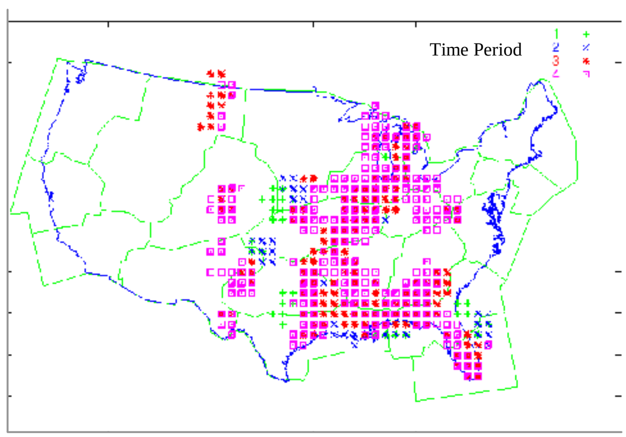

Figure 7. Gridded Severe Weather Polygons, 6/11/99
(Symbols indicates time—1: 6am-10am, 2: 10am-2pm, 3:2pm-6pm, 4: 6pm-11pm Eastern Time)

a distance measure such as Euclidean (see [4] for an introduction to the topic). Initial trials using conventional distance measures in clustering software packages were not successful. Plots of days declared similar by this approach were not visually appealing. It was thought that the conventional distance measures were inappropriate for the data, which consisted of very many zeroes (i.e., no severe weather in the cell during that time period). In order to account for this situation, a new distance measure was formulated. A distance measure was created that declared *closeness* as a function of cell-specific agreement (the 2 days being compared both had a given cell valued non-zero, i.e., both experienced severe weather), but declared "distance" if there was cell-specific disagreement (the 2 days being compared had a given cell with differing senses: either zero/non-zero or non-zero/zero). Presenting the new computed distance matrix to the clustering algorithm yielded acceptable results—sufficiently similar days were grouped together to create day-types.

-

4.2 Creating Scheduled Traffic Day-Types

In an analogous fashion, day-types were created for scheduled traffic. For each subject day, the top 50 origin-destination pairs with respect to scheduled flights were considered. Flight frequency was mapped into one-part-in-three, and used as a weighting scheme.

4.3 Model Building

The two day-type schemas were used in regression modeling. The basic assertion of the modeling effort is that, at a high level, system performance is a function of traffic, weather, and other factors:

NAS Impact = f (Traffic Level and Patterns, Weather, Other Factors)

If some influence of interest can be represented as the third explanatory variable, the influence and interaction of causal factors can be examined. As described earlier, the response variable is a composite performance measure, the FIV, which is the percentage of flights with departure delays exceeding 30 minutes plus the percentage of cancellations plus twice the percentage of diversions.

As a trial usage of the proposed modeling, a specific major U.S. carrier was considered. This carrier, which will go unnamed, is designated here as *ABC*.

The regression model is, then:

FIV = f (Some Combination of Explanatory Variables: **TrfDay**, **WxDay**, **Season**)

where:

TrfDay is a categorical variable indicating day-type with respect to scheduled traffic level and pattern

WxDay is a categorical variable indicating day-type with respect to severe weather location and extent

Season is Spring or Summer 1999 or 2000

Note that the variables FIV and TrfDay are germane to the specific carrier ABC. The WxDay variable is not airline-specific. The third explanatory variable, Season, is actually rather coarse—little is known about the ABC airline operations. Company-internal phenomena such as staffing, new scheduling tools or initiatives, or new market development are not represented.

4.4 Regression Analysis

Using the above relationship, various regression analyses with differing sets of explanatory variables were run. The influence of weather, traffic, and season as single regressors and in combination with each other was assessed. Table 1 shows these results. The left-most column shows a model number referenced in the discussion below.

Table 1. Regression Modeling Results

Model No.	Explanatory Variables				Model p (small is good)	Adj. R^2 Whole Effect (large is good)	Adj. R^2 Best Model (large is good)
1	TrfDay				<0.0001	0.38	0.42
2		WxDay			<0.0001	0.43	0.51
3			Season		<0.0001	0.20	0.20
4	TrfDay	WxDay			<0.0001	0.54	0.63
5	TrfDay	WxDay		TrfDay*WxDay	<0.0001	0.67	0.77
6	TrfDay	WxDay	Season		<0.0001	0.55	0.63
7	TrfDay	WxDay	Season	TrfDay*WxDay	<0.0001	0.66	0.77

An interpretation of Table 1 is possible, using the columns labeled "Model p" for observed significance of the model, and Adjusted R-squared. Two columns of Adjusted R-squared values are shown, with sub-labels *Whole Effect* and *Best Model*. Whole Effect refers to a model using all of the variables, whereas Best Model uses the best (highest adjusted R^2) subset of

-

contrasts to extract the most information. The following can be observed:

- As lone independent variables, traffic, weather, and season each explain some amount of the variation in the response variable. Weather explains the most, and season explains the least (see Model Nos. 1, 2, and 3). (Note that the differences in mean response attributable to Season [not shown] were as expected: Spring and Summer 1999 were low and about equal, Spring 2000 was higher, and Summer 2000 was higher still.)
- There is some evidence for inclusion of a term in a combined model to represent the interaction effect of traffic and weather (Model Nos. 4 vs. 5).
- In a combined model, Season adds no explanatory power given traffic and weather (see Model Nos. 4 vs. 6 and 5 vs. 7). The stepwise regression procedure [not shown] suggests that Season not be included in a final combined model.

5 Conclusions

To restate, the study objective is to seek a means of normalizing for weather effects, allowing a fairer comparison of the daily performance of the NAS under varying conditions. This objective was explored using two methods. The first method computes a continuously-valued scalar index (i.e, WITI) for a day. The second method calculates *day-types* or discrete groupings of day which are similar with respect to severe en route weather. Using statistical measures such as correlation coefficients and Smirnov statistics as the figures of merit, both methods yield reasonably significant results.

This study represents the first phase, the basic research, of a two-phase project. Both methods demonstrate potential, and should be taken forward to a second phase which would explore practical applications. Already both methods have been used to investigate whether there were significant differences between Spring/Summer 1999 and Spring/Summer 2000 in terms of certain measures of air traffic system performance after adjusting for the effects of weather; no significant difference was found.

6 Acknowledgments

The work reported here builds on work performed in fiscal year 2000 for the Federal Aviation Administration (FAA) under contract DTFA01-93-C-00001. The authors acknowledge helpful discussions with Mr. Rich Jehlen (then ATP-420) and Mr. Steve Bradford (ASD-130) of the FAA during that work and the work reported here; many of the ideas for en route WITIs resulted from those discussions.

7 References

[1] NAS Operational Evaluation Plan Version 3.0, Federal Aviation Administration, June 2001

[2] Information Theory and an Extension of the Maximum Likelihood Principle - H. Akaike Second International Symposium on Information Theory - Budapest - B.N. Petrov and F. Csaki, editors - Akademiai Kaidó - pp. 267-281 - 1973

[3] R: A Language for Data Analysis and Graphics - Ross Ihaka and Robert Gentleman - Journal of Computational and Graphical Statistics - vol. 5, no. 3, pp. 299-314 - 1996

[4] Classification: Methods for the Exploratory Analysis of Multivariate Data - A.E. Gordon - Chapman and Hall - New York, NY, USA - 1981

Revised 2001-12-24.

8 Author Biographies

Dr. Michael Callaham is a Senior Operations Research Analyst at The MITRE Corporation, where he has been employed since 1995. Dr. Callaham holds an Sc.D. degree in Instrumentation, an S.M. degree in Aeronautics and Astronautics, and an S.B. degree in Mathematics, all from M.I.T. He is a Senior Member of the American Institute of Aeronautics and Astronautics and a Member of its Modeling and Simulation Technical Committee, a Member of the Institute of Electrical and Electronics Engineers, and a Member of the International Institute of Forecasters. At MITRE, he uses air traffic simulation to forecast benefits of proposed improvements in air traffic management systems and procedures.

Jim DeArmon is a Lead Engineer with The MITRE Corporation. He has worked for the last several years in research and modeling of air traffic flow dynamics. He holds a M.S. degree in Operations Research, and has taught graduate courses at the University of Maryland. He has written numerous papers in applied math and aviation.

Arlene M. Cooper is a Senior Economics/Business Analyst with The MITRE Corporation's Center for Advanced Aviation System Development (CAASD). She has experience in process modeling, operational analysis and economic impact analysis in support of various governmental agencies. Ms. Cooper has an M.S. degree in Operations Research from George Mason University and a B.S. degree in Mathematics from Georgetown University.

Dr. Jason H. Goodfriend is now a systems engineer at The MYTA Corporation and is currently modeling the performance of proposed Federal Aviation Administration (FAA) systems. He has worked on a variety of projects for the Federal Aviation Administration. Past work involved determining quantitative measure of performances for the achievement of FAA goals. In addition, he has been involved with statistical accuracy studies for the Global Positioning System (GPS) and analysis of communication protocols for the Aeronautical Telecommunication Network (ATN). Dr. Goodfriend was employed by The MITRE Corporation when contributing to this paper and, before that, by Stanford Telecommunications. He holds a Ph.D. degree in Systems Engineering and an M.S. degree in Mathematics from the University of Virginia.

Debra Moch-Mooney is a Senior Multi-Disciplinary Systems Development Engineer with The MITRE Corporation's Center for Advanced Aviation System Development (CAASD). She has worked the last several years on projects using models to analyze and improve the operational performance of airspace and airports. She holds a M.S. degree in Operations Research from the University of Pennsylvania.

Mr. George H. Solomos is a Project Team Manager at The MITRE Corporation's Center for Advanced Aviation System Development (CAASD). Mr. Solomos graduated Magna Cum Laude with a B.S. degree in Computer Science from the University of Maryland Baltimore County in 1991. In 1996 he received a M.S. degree in Operations Research from The George Washington University.

Mr. Solomos' work for The MITRE Corporation began in 1990 when he worked on developing a computer human interface prototype for the Advance AERA (Automated En Route Air Traffic Control) Concepts project. Over the years at CAASD Mr. Solomos has worked extensively with the Federal Aviation Administration (FAA), especially in the traffic flow management domain. He has performed benefit analyses and system performance forecasting, completed simulation model verification and validation, developed software, explored data visualization and data analysis techniques, and explored issues related to changes in oceanic procedures for the FAA's Oceanic Operations and Procedures branch. Mr. Solomos has performed in-depth analysis at several Air Route Traffic Control Center (ARTCC) and Terminal Radar Approach Control (TRACON) facilities and he spent several years working on near term procedural enhancements with the Air Traffic Control System Command Center.

Currently Mr. Solomos is working in the Air Transportation Performance and Policy Analysis program area in CAASD. He is focusing on post-implementation assessments, future performance estimation, system performance analysis and air carrier service modeling.