

DECISION SUPPORT TOOL FOR PREDICTING GROUND DELAY PROGRAMS
(GDP) AND AIRPORT DELAYS FROM WEATHER FORECAST DATA

David Smith (PhD Candidate)

Email: dsmit4@gmu.edu

Phone: (703) 975-1386

Lance Sherry (PhD)

Email: lsherry@gmu.edu

Phone: (703) 993-1711

Fax: 703-993-1521

Center of Air Transportation and Systems Research

Department of Systems Engineering and Operations Research

George Mason University

4400 University Dr.

Fairfax VA 22030

ABSTRACT

During peak periods of operation, the National Airspace System (NAS) suffers from limited resources. Traffic Flow Management (TFM) is responsible for adjusting traffic flow demand to meet the available capacity. In the case of airports, TFM uses the Ground Delay Program (GDP) to allocate the constrained runway resources.

This paper describes a tool for forecasting the likelihood of a GDP at an airport and the delays at an airport throughout the day and at specific time periods during the day. The tool, trained on historical data, predicted GDP's at four OEP airports with accuracy at 73% and delays with accuracy of 76%. The implications of these results are discussed for airlines and TFM.

INTRODUCTION

For efficient operation of the NAS, there is a need for the weather forecasting services and TFM products to estimate the reduction in capacity due to adverse weather. Weather forecast products are uncertain and the uncertainty increases with lead-time. Useful applications of weather forecasts requires either refinement, consultation, and application of the weather forecast to estimate air traffic capacity or decision support tools that take forecasts and make predictions based on past forecasts and those forecasts connections to NAS capacity.¹ This paper describes a methodology used to create one such decision support tool known as the Weather Delay Prediction Tool. With this tool, the user enters the TAF for a given day and airport and the tool provides a GDP prediction, delay predictions, and estimated AARs.

While weather forecasts are not 100% accurate, meteorologists have the ability to predict weather with increasing accuracy. Through a complex system of weather services, government agencies, pilots, and other aviation professionals receive the vast knowledge base in the form of up to date weather forecasts. These forecasts enable air traffic managers to make informed decisions regarding weather and flight safety. Most weather related research has focused on the improvement of the accuracy of the forecasts and for the pure weather professionals, this is worthwhile research. For the air traffic professional however, research should be focused on how the forecast affects the actual NAS performance.^{2 3} Is there a potential for a Ground Delay Program (GDP)? What will be the average delay? What will be the aircraft arrival rates? It is not whether it rains which is important, but instead did the forecast make you bring your umbrella?

These questions can be answered by developing a model using a pattern recognition tool known as a support vector machine (SVM). The SVM is trained with past data and then develops functions to enter present day data to predict future outcomes. The SVM is superior to other techniques such as regression because it does not require a linear relationship between the data. This way we can use individual data points with few possible choices and still come up with predictive functions. As seen in Figure 1, we collect TAF data, apply the SVM, produced GDP and delay predictions, and then apply queuing theory to predict AARs.

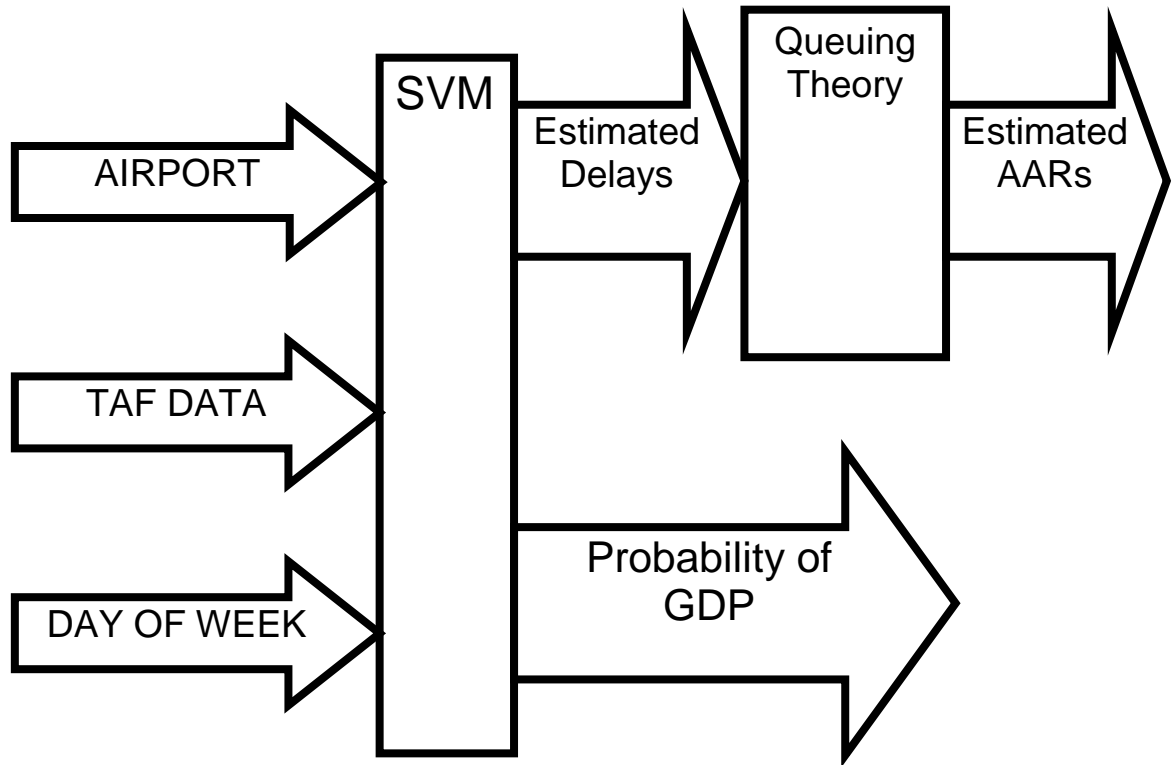


FIGURE 1 SVM Method

DATA AND METHODOLOGY

The Terminal Aerodrome Forecast

A variety of different forecasts products are produced and designed to be used in the preflight planning stage. One of these forecasts that traffic managers should be familiar with is the Terminal Aerodrome Forecast (TAF). A TAF is a report established for the 5 statute mile radius around an airport. Each TAF is valid for a 24-hour time period and is updated four times a day at 0000Z, 0600Z, 1200Z, and 1800Z. The TAF is reported in a standard international format code and must follow a set of rules that define what must be placed in each line and what criterion requires a new line. There are four different lines in a TAF, the first one gives location, valid time and given weather for that time until the next line of the forecast. The BECMG (becoming) line indicates between the two hours given, the weather will start to change from the previous line to the next line; while a FM (from) line indicates that after the given time, the weather will be what the line states; and the TEMPO line indicates that for time period given, the weather parameters in the line will happen for no more than 30 minutes of the hour, but will happen for at least 1 minute of the hour. This code may seem confusing to the casual viewer, however, this code allows a computer program to easily search for and identify certain weather conditions and the time that these conditions are forecasted.⁴

For this paper we collected all available archived TAF data for Newark, O'Hare, Atlanta, and Philadelphia airports. This data is only available from January 2002 to April 2007. Each set of TAF data was loaded into an Excel spreadsheet. Within the spreadsheet we programmed a macro to find, for four time periods during the day, the wind speed and direction, visibility, and ceiling conditions. For the same time periods,

we also extracted binary variables to indicate whether rain, snow, thunderstorms, fog, showers, and or mist was forecasted. Finally, by using the airport runway chart, we determined whether the wind was a cross wind and entered it as a binary variable. We choose forecast time periods of 1100Z, 1500Z, 1900Z, and 2300Z because these 4 hour intervals corresponded to peak airport operating hours between 0600 to 2200 local time. These variables created a vector of 56 values in length. The vector was extended to 57 to add a value corresponding to the last digit of the year as we found that delays and GDPs were on the increase each year. Using these vectors developed from past forecasts, we intend develop functions using the Support Vector Machines (SVM) to predict future air traffic management decisions and outcomes.

Support Vector Machines

Support Vector Machines are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. A schematic example is shown in the Figure 2.

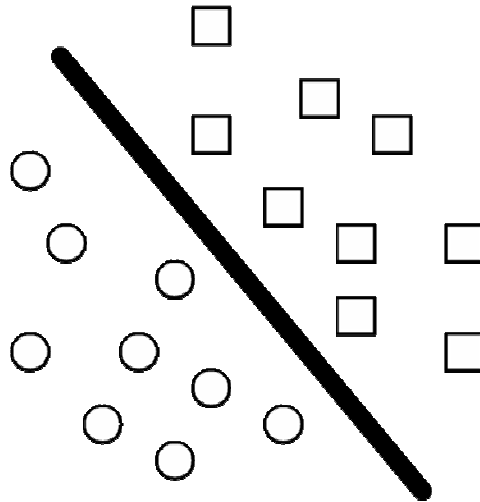


FIGURE 2 Separating line defines a boundary.

In this example, the objects belong either to class square or circle. The separating line defines a boundary on the right side of which all objects are squares and to the left of which all objects are circles.

Figure 2 is a classic example of a linear classifier, i.e., a classifier that separates a set of objects into their respective groups (square and circle in this case) with a line. Most classification tasks, however, are not that simple, and often more complex structures are needed in order to make an optimal separation, i.e., correctly classify new objects (test cases) on the basis of the examples that are available (training cases). This situation is depicted in Figure 3.

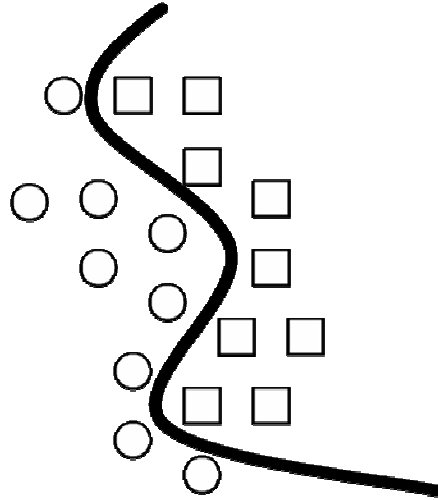


FIGURE 3 Full Separation requires a curve.

Compared to Figure 2, it is clear that a full separation of the square and circle objects would require a curve (which is more complex than a line). Classification tasks based on drawing separating lines to distinguish between objects of different class memberships are known as hyper-plane classifiers. Support Vector Machines are particularly suited to handle such tasks.

Figure 4 shows the basic idea behind Support Vector Machines.

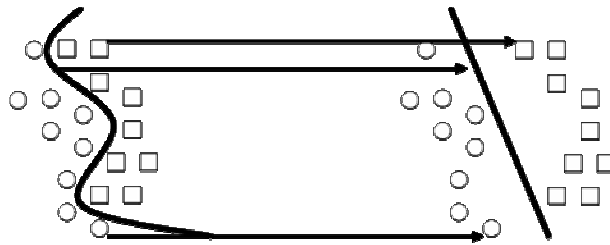


FIGURE 4 Objects are mapped using a set of mathematical functions.

Here we see the original objects (left side of the schematic) mapped, i.e., rearranged, using a set of mathematical functions, known as kernels. The process of rearranging the objects is known as mapping (transformation). Note that in this new setting, the mapped objects (right side of the schematic) is linearly separable and, thus, instead of constructing the complex curve (left schematic), all we have to do is to find an optimal line that can separate the square and the circle objects.

Support Vector Machine (SVM) is a method that performs classification tasks by constructing hyper-planes in a multidimensional space that separates cases of different class labels. SVM supports both regression and classification tasks and can handle multiple continuous and categorical variables. For categorical variables a dummy variable is created with case values as either -1 or 1. For this type of SVM, training involves the minimization of the error function:

$$\begin{aligned} & \frac{1}{2} w^T w + C \sum_{i=1}^N \xi_i \\ \text{s.t. } & y_i (w^T x + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, i = 1 \dots N \end{aligned}$$

Where C is the capacity constant, x is the vector of coefficients, b a constant, y the dummy variable, and ξ_i are parameters for handling non-separable data (inputs). The index i labels the N training cases. The kernel w is used to transform data from the input (independent) to the feature space. It should be noted that the larger the C the more the error is penalized. Thus, C should be chosen with care to avoid over fitting.⁵

DATA COLLECTION

For this paper we collected data for the TAF, delay statistics and GDP.

Terminal Aerodrome Forecast

The TAF data was collected from a website provided by the National Climatic Data Center. These files tend to be long, up to 100 pages of text data, because all reports received are placed in these files as they are received and they are updated approximately every five minutes as data becomes available. Also, forecasts may be duplicated within the files and multiple forecasts received from a station may appear in a file.⁶

Ground Delay Program Data

The FAA publishes summaries for Ground Delay Programs within its Collaborative Decision Making (CDM) website. These summaries list the significant CDM events (e.g., GDP parameters, compression times, program revisions, etc.) that would assist in the re-creation of a particular program using FSM Historical Data Mode. The information in these summaries was culled from various manually recorded logs, including the GDP digest recorded by the air traffic management specialist running the GDP. For our purposes we only used the airport, date, and time data for each GDP.⁷

Delay Data

The final data collected was the delay statistics for individual airports. We found this data on the Bureau of Transportation Statistics website for summary statistics for destination airports. From the website we collected Total Number of Flights, Arrival Delay (minutes), and Arrival Time for each day. With this information, we then found the average daily delay and the delay for the time periods 0600 to 1000, 1000 to 1400, 1400 to 1800, and 1800 to 2200 local time.⁸

METHODOLOGY

This paper used the SVM method to determine GDPs and delays, while we applied queuing theory to determine AARs.

Method for Predicting GDP

After collecting the TAF data as the independent variable matrix and the GDP data as the dependent variable vector we applied the SVM to determine a function to predict future

GDP's. The quadratic program introduced earlier was coded into AMPL. AMPL is a comprehensive and powerful algebraic modeling language for linear and nonlinear optimization problems, in discrete or continuous variables. After coding, we submitted the program and the associated data to the NEOS Server for Optimization.

In our specific case, the x in the quadratic program represented the 57 character long vector from the TAF weather data collected from 2002 through 2006. The wind speed and direction data and the ceiling heights were all the actual values found from the TAF data while binary variables were used to indicate whether or not certain weather conditions were predicted. For our case the y represents a binary variable that indicates whether or not a GDP was established for a given airport. Values equal to -1 indicate no GDP while values equal to 1 indicate that there was a GDP. The C variable was set at 1000 after experimenting with other values. The ξ_i vector was determined, but was only used to relax the function, so a solution was possible. It was not used in the final prediction function.

For the independent variables, the five years worth of data included 1826 days so this created an 1826 by 57 data matrix for the independent variable. Then, for all 4 airports in the study, we created a y vector that contained 1 if there was a GDP that day and -1 if there was not a GDP that day. We ran the AMPL code on the NEOS Server and found a solution vector w and variable b for each airport. With the determined vector and variable we could then look at a current TAF forecast and develop an x vector using that data and then use the equation below to develop a prediction value.

$$w^T x_i + b$$

If the prediction value was greater than 0, then we predict that a GDP will occur on that day. Conversely, if the value is less than zero then we predict a GDP will not occur on that day.

Method for Predicting Delay

The method for predicting delay is similar to the method for predicting GDP. However, the dependent data for the GDP was merely binary while the delay data was not. To account for this difference we had to do several runs of the SVM for each airport. Instead of dividing the data based on whether or not there will be a delay we divided it based on minute interval bins. For each airport we divided the delay data into four bins with various minute intervals. For example, Newark had bins at 15, 30, 45, 60 minute intervals. For this SVM operation, the y vector for the 15 minute delay contained a -1 if the delay was less than 15 minutes and a 1 if the delay was or exceeded 15 minutes. For the 15 minute delay interval there were more ones than negative ones, but as the time interval increased the number of positive ones decreased and the number of negative ones increased. This created four predictor functions for each airport.

To predict the delay for a present day forecast we had to apply the 4 functions to the x vector for each day in the same way as we did for predicting GDP. Ideally you would have a series of ones followed by a series of negative ones. The first negative one encountered would indicate that the time interval before that was the predicted delay. For instance, if you found a -1 under the 30 time second interval, then the predicted delay for that particular day would be 15 minutes.

Newark had the largest average daily delay of any of the airports in our study and the 15 minute interval to 60 minutes worked. The other airports in the study did not have as large daily delay and had few delays that exceeded 45 or 60 minutes. The SVM was not as effective at recognizing any of these longer delays, so the intervals were shortened.

Method for Determining Aircraft Arrival Rates

To predict the arrival rates, we used queuing theory and assumed an M/M/1 queuing system. The M/M/1 queuing system has exponential inter-arrival times in which we assume that the arrival rate per unit time is λ and a single server with exponential service times in which we assume that each customer's service time is exponential with rate μ .⁹ We assume that the delay is the average time a customer spends in the system, so the delay can be determined using the equation below.

$$D = \frac{1}{\mu - \lambda}$$

The arrival rate per unit time λ is estimated by using the total number of flights that arrived for each particular day and divide that by 16 operating hours. Since the predictor model determines the delay, we can use algebra to determine customer's service rate μ which is assumed to be the aircraft arrival rate by using the equation below:

$$\mu = \frac{1}{D} + \lambda$$

RESULTS

Weather Delay Prediction Tool

Using what we have learned from the SVM method, we can create a tool that will predict future average daily delay, average delay for certain time periods, and whether or not a GDP will be used. The TAF is loaded into an Excel spreadsheet and then a macro is applied to extract the required data for each day which is the x vector. Then we use the w vector and b variable found using the SVM to predict the average daily delay and GDP using the procedure describe in METHODOLOGY section. From the delay data, we found that by using linear regression, you can develop a linear equation to create a reasonable estimate of the average delay at given time periods. With the knowledge of the delay and the length of these time periods we can use the queuing theory described earlier to come up with AARs for each of these periods. The values are displayed on an Excel worksheet.

Accuracy in Predicting Ground Delay Programs

A prediction is valid if it detects most airports with a GDP and excludes most airports without a GDP, and if a positive prediction usually indicates that a GDP will be used as a traffic flow management initiative. To evaluate whether or not the SVM technique is effective we computed the sensitivity, specificity, positive predictive value and the negative predictive value of the technique. Sensitivity is the proportion of truly non-GDP days that are identified as GDP days by the SVM technique. Sensitivity is a reflection of

the true positive rate. In other words, of all those GDP days how many are picked up by the SVM technique? Specificity is the proportion of actual non-GDP days that are so identified by SVM technique. Specificity is a reflection of the true negative rate. In other words, of all those days that are truly non-GDP days, how many are picked up as non-GDP airports by the technique? The positive predictive value (PPV) is the probability that if the SVM technique indicates a GDP at an airport, that a GDP will actually occur. The negative predictive value (NPV) is the probability that if the SVM technique indicates no GDP, that a GDP will not be used.

These values were computed by training the SVM using 2002 through 2006 data to find the w vector and b variable. We then applied predictor function back onto the training data to see how well it worked. The values used to evaluate the prediction are found in Table 1.

Airport	Sensitivity	Specificity	PPV	NPV	% Correct
EWR	0.48	0.91	0.74	0.77	0.76
ORD	0.35	0.91	0.67	0.72	0.72
ATL	0.45	0.90	0.63	0.81	0.78
PHL	0.46	0.94	0.69	0.86	0.84
Average					0.78

TABLE 1 Prediction Compared to Actual Values

From Table 1, you can see that the SVM method predicts whether or not a GDP will be at an airport an average of 78% of the time. Table 1 also shows that the model is more effective on predicting that there will not be a GDP. This is because non-GDP days are the norm. In Table 1, we predicted the past, so then we had to see how well the results fit the future. In Table 2 we used the model with the first two months of 2007.

Airport	Sensitivity	Specificity	PPV	NPV	% Correct
EWR	0.66	0.67	0.78	0.52	0.66
ORD	0.23	0.81	0.42	0.63	0.58
ATL	0.67	0.89	0.25	0.98	0.88
PHL	0.53	0.86	0.57	0.84	0.78
Average					0.73

TABLE 2 Predictions Compared to Future Values

Unfortunately, we were able to only collect the January and February GDP's for 2007, so the data set is small. However, there are still some encouraging results with sensitivity values at EWR, ATL, and PHL all exceeding 0.53 and specificity values exceeding 0.67 for all airports in the study. Also the model predicts GDPs accurately 73% of the time. A larger data set may provide even better results.

Delays

Earlier we developed a function for minute delay intervals for each of the four airports in the study. As with GDP, we trained the SVM using TAF data from 2002 to 2006. We then took the function and applied it back to the training data to determine the fit. The results from the four airports can be found in Table 3.

You can see from Table 3 that the SVM model predicts whether or not there will be a delay exceeding the given interval at least 72% of the time and the accuracy can be as high as 97% with an average of 83% of the time. This shows that the function is a reasonable fit for the data. Table 4 uses the first four months of 2007 as test data to see how well the model works on future data.

Airport	Average Delay	Sensitivity	Specificity	PPV	NPV	% Correct
EWR	15 min	0.46	0.94	0.78	0.79	0.79
EWR	30 min	0.40	0.97	0.75	0.89	0.88
EWR	45 min	0.22	0.99	0.78	0.94	0.93
EWR	60 min	0.18	0.99	0.97	0.98	0.97
ORD	7.5 min	0.35	0.92	0.69	0.72	0.72
ORD	15 min	0.27	0.96	0.68	0.80	0.79
ORD	22.5 min	0.25	0.98	0.74	0.85	0.84
ORD	30 min	0.12	0.99	0.76	0.88	0.88
ATL	5 min	0.58	0.87	0.78	0.72	0.74
ATL	10 min	0.52	0.89	0.69	0.81	0.78
ATL	15 min	0.34	0.93	0.62	0.83	0.81
ATL	20 min	0.19	0.97	0.58	0.97	0.88
PHL	5 min	0.53	0.89	0.77	0.74	0.74
PHL	10 min	0.47	0.94	0.75	0.83	0.81
PHL	15 min	0.48	0.96	0.74	0.88	0.86
PHL	20 min	0.36	0.98	0.73	0.90	0.88
Average						0.83

TABLE 3 Airport Results from Training Data

Airport	Average Delay	Sensitivity	Specificity	PPV	NPV	% Correct
EWR	15 min	0.65	0.86	0.87	0.65	0.74
EWR	30 min	0.53	0.88	0.73	0.76	0.75
EWR	45 min	0.35	0.98	0.78	0.88	0.88
EWR	60 min	0.00	0.95	0.00	0.91	0.87
ORD	7.5 min	0.33	0.98	0.96	0.48	0.58
ORD	15 min	0.35	0.94	0.81	0.68	0.70
ORD	22.5 min	0.16	0.99	0.99	0.68	0.70
ORD	30 min	0.06	0.99	0.99	0.74	0.74
ATL	5 min	0.35	0.93	0.76	0.70	0.71
ATL	10 min	0.32	0.98	0.82	0.83	0.83
ATL	15 min	0.24	0.98	0.67	0.89	0.88
ATL	20 min	0.25	0.97	0.50	0.92	0.90
PHL	5 min	0.53	0.83	0.83	0.52	0.64
PHL	10 min	0.34	0.92	0.79	0.61	0.65
PHL	15 min	0.53	0.93	0.77	0.81	0.80
PHL	20 min	0.48	0.95	0.72	0.86	0.84
Average						0.76

TABLE 4 Airport Results from Test Data

Table 4 shows that the SVM method continues to predict delays times with an average of 76% accuracy.

CONCLUSIONS

Recommendations

We conclude from these results that it is possible to create a tool that inputs the current TAF forecast and produces delays predictions hours in advance. There are tool major barriers to air traffic flow, scheduled congestion and weather. Scheduled congestion can be predicted by using simulation tools that fly the current commercial airline schedules.

The Weather Delay Prediction Tool helps with the weather part by providing the time and place of weather delays. Given this information, air traffic managers can create branch plans to counter any potential delays. They can then publish these plans so the airlines know what may happen and what actions are going to be taken if congestion does occur. This increased predictability should increase NAS efficiency.

Future Work

The NAS is a complex and inter connected system where delays at one airport may cause delays at another. Therefore, further study should be focused on how TAF reports affect airports other than the one the forecast was created for. Also sensitivity analysis should focus on what how many different airport forecasts should be included to create a delay prediction at a specific airport. Finally, is there enough connectivity in the system to predict delays at airports based on TAF forecasts from other airports close by and not using the TAF report for the home airport? The smaller the set of TAF reports required, the simpler the TAF prediction tool needs to be.

ACKNOWLEDGEMENT

We would like to thank the Center for Air Transportation System Research (CATSR) at George Mason University for providing support to this study. We would also like to thank Dr. George Donohue from George Mason University and Dr. David Rodenhuis from the University of Victoria, British Columbia. Earlier conversations with both were the genesis of this work and their advice on air traffic management and weather forecasting were vital. We would also like to thank Ved Sud from the FAA, Mike Brennan from Metron Aviation, and Mike Wambsgans from Flight Explorer for their technical contributions. Danyi Wang and Bengi Manley from the CATSR lab provided valuable information on delays and ground delay programs and provided input to early study results.

¹ Rodenhuis, Dave, "A Concept of Operations for an Interactive Weather Briefing", Federal Aviation Administration. Herndon, Virginia, 2006.

² Torbert, Duane and Dave Rodenhuis, "Operational Feedback Reports to Providers of Aviation's Collaborative Convective Forecast Product", Federal Aviation Administration. Herndon, Virginia, 2005.

³ Kay, Michael P., Jennifer L. Mahoney, and Joan E. Hart, "An Analysis of CCFP Forecast Performance for the 2005 Convective Season", University of Colorado, Bolder, CO, 2006.

⁴ *Pilot's Handbook of Aeronautical Knowledge*. Federal Aviation Administration. Washington, DC, 2003.

⁵ Hill, T. Lewicki, P., *STATISTICS Methods and Applications*. StatSoft, Tulsa, OK, 2006.

⁶ HDSS Access System, National Climatic Data Center, U.S. Department of Commerce, Washington, DC. Hurricane.ncdc.noaa.gov. Accessed June 2007.

⁷ GDP Summaries, Collaborative Decision Making, Federal Aviation Administration, Washington, DC. Cdm.fly.faa.gov. Accessed June 2007.

⁸ TransStats, Bureau of Transportation Statistics, Department of Transportation, Washington, DC.

⁹ www.transtats.bts.gov. Accessed June 2007.

⁹ Winston, W.L., *Operations Research: Applications and Algorithm*. Brooks/Cole – Thomas Learning, Belmont, CA, 2004.