Eighth USA/Europe Air Traffic Management Research and Development Seminar (ATM2009)

Modeling Flight Delays and Cancellations at the National, Regional and Airport Levels in the United States

Banavar Sridhar, Yao Wang
NASA Ames Research Center
Moffett Field, CA, USA

Alexander Klein
Air Traffic Analysis, Inc.
Fairfax, VA, USA

Richard Jehlen
The U.S. Federal Aviation Administration
Washington, DC, USA

Abstract - **This paper describes models for predicting weather-related aircraft delays and cancellations at the national, regional and airport levels. The models estimate delay based on the number of aircraft affected by expected weather. The estimation and prediction models incorporate both regression methods and neural networks, using two operational databases. The paper compares the performance of traditional linear regression models with several neural network models in the estimation of key airspace metrics such as total aggregate delay, arrival delay, airborne delay, and flight cancellations. Conclusions are: (a) the metric based on the number and extent of aircraft expected to be impacted by weather is a good proxy for delay at all levels, (b) different delay models are preferable for different seasons and the delay estimation accuracy is higher in the convective weather season than the non-convective weather season, (c) the delay estimation accuracy at all levels and for different metrics is about the same, (d) models resulting from using either database are complementary and provide similar accuracy and (e) the neural network delay models perform slightly better in that they have a higher correlation between model output and airspace metrics than linear regression methods.**

Keywords – **Delay, Weather impact, Metrics, National Airspace System models, Neural networks**

I. INTRODUCTION

Of all airspace delays occurring, weather is the most significant causal factor by a wide margin [1]. To guide appropriate traffic flow management decisions, it would be useful to first have a delay baseline and then be able to model the relationship between weather and delays above the baseline.

Efforts have been made to understand the connection between weather and delay at the local and national levels [2-8]. A common theme is to use of the Weather Impacted Traffic Index (WITI), which measures the number of aircraft affected by weather at a given time. Delay models have been developed using linear and non-linear regression models, and single performance metrics have originated form either one of two FAA data sets. This paper provides a comprehensive study of multiple performance metrics drawn from both data sets at the national, regional and airport level. Another contribution of this paper is a comparison of airspace performance modeling using linear regression and neural network approaches.

In this paper, Section 2 provides a review of airspace performance metrics. Different definitions of WITI are discussed in Section 3. Section 4 describes the algorithms used for model development. In addition, the section provides a neural network modeling overview and discusses factors affecting the training, error performance and their validation. Section 5 presents results from using linear regression methods and neural networks for delay estimation and cancellations at the national level. Airport delay models are presented in Section 6. Finally, conclusions are provided in Section 7.

II. NAS PERFORMANCE METRICS AND DATABASES

Two FAA data sources, OPSNET and ASPM, include measures of air traffic control system performance. OPSNET collects data reflecting delays resulting from FAA actions to maintain a safe system in the presence of congestion and bad weather. The OPSNET data are available starting from 1990, whereas the ASPM data are relatively new starting from the year 2000 [9]. The OPSNET delays are counted when a flight is delayed more than 15 minutes compared to the flight plan time filed with the FAA. The OPSNET total delay is the sum of all delays experienced by all aircraft

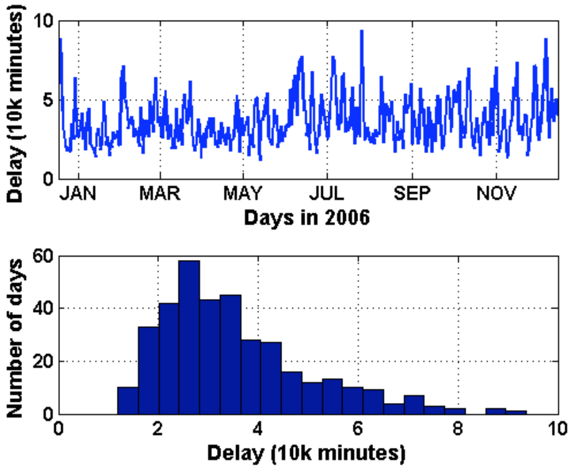

Figure 1. Arrival delay and its histogram for 2006.

during a day. The OPSNET delays are further categorized by flight category, class of traffic and cause of delay. The ASPM delays are measured relative to flight plan filed with the FAA or to air carrier schedules from the OAG and carrier computerized reservation systems. The OAG [10] is best known for its world-wide airline schedules database.

The delays are based on flight data from the Enhanced Traffic Management System (ETMS) and other information sources (e.g., airline schedules, operations and delays, weather information, runway information, etc.). ASPM delays are a measure of actual delays experienced by the airlines and its customers. ASPM collects data at a finer granularity, reports delays of one minute or more and classifies delay by all phases of flight and time of the day. The ASPM also provides the daily number of flights cancelled by the airlines. Both systems contain data entered by human operators and are prone to data recording errors. The two systems are overlapping in certain areas and complementary in others. Both databases can be used independently for developing NAS metrics models based on statistical analysis.

Table 1. Characteristics of two FAA operational databases

	OPSNET delay	**ASPM delay and cancellation**
Data range	1990-present	2000-present
Sample rate	Daily values	Quarter hour values Daily value: sum of quarter
Airports	National values for all airports, Centers, etc.	71 Airports National values: sum of all airports
Perfor-mance metrics	Total delay time	Total OAG-based (schedule), flight plan based arrival delay time, EDCT hold minutes, airborne delay times, and flight cancellation

Table 1 shows the basic characteristics of OPSNET and ASPM data; Table 2 provides a statistical description of daily metrics based on 1096 days in 2004, 2005 and 2006. In Table 2, all the metrics except OPSNET total delay are from ASPM. The variation of all metrics in Table 2 excluding OAG-based arrival delay for the years 2004, 2005 and 2006 are shown in Figure 1, top and their histogram, in Figure 1, bottom. The large difference between OPSNET and ASPM delays are due to OPSNET delays accounting only delays greater than 15 minutes and the result of a FAA decision to manage traffic safely.

Table 2. NAS performance metrics for 2004-2006

Daily delays in 1000 minutes	**Median**	**Mean**	**Absolute deviation**
OPSNET total delay	46	58	26
Flight plan based arrival delay	211	230	60
OAG-based arrival delay	259	289	75
OAG-based (>=15) arrival delay	229	259	72
EDCT hold minutes for airports	36	47	26
Airborne delay	82	81	26
Count of arrival cancellations	318	410	186

The Pearson correlation coefficient r, $0 \le r \le 1.0$, [11] is used to compare the dependence between two variables. A correlation greater than .8 is generally considered as strong whereas a correlation of less than .5 is generally treated as weak. As an example, if r = .80, then r^2 = .64, which means that 64% of the total variation in the output can be explained by the linear relationship between the variables. Table 3 shows correlations between OPSNET and ASPM metrics. OPSNET total time delays have good correlations with ASPM flight plan based and OAG-based delays. OPSNET and ASPM EDCT delays are very similar; the difference could be due to the fact of that ASPM delays do not include flights held by Centers.

Table 3. Correlation between OPSNET and ASPM delay metrics

OPSNET delay vs. ASPM delay Correlation	***r***
Total OAG-based arrival (Schedule)	.83
Total OAG-based arrival (>= 15 minutes)	.83
Total flight plan-based arrival minutes	.84
Total EDCT hold minutes by airports	.91
Total airborne delay minutes	.55
Total count of cancellations	.38

III. WITI AND ITS VARIATIONS

Over the past several years, a number of studies have been conducted to understand the connection between weather and delay. The motivation for the creation of an air transportation weather index was to establish a consistent, quantified measure reflecting the impacts due to weather occurrences across the air transportation network. For example, a weather phenomenon between the New York and Chicago metropolitan areas imposes substantially more impact on the air transportation system than the same phenomenon (severity, duration, scope, etc) occurring between Bismarck and Bozeman. Studies using the concept of WITI [1-6] have established that its aggregate national weather index has strong correlation with national OPSNET delays.

Two different definitions of WITI are used in this paper to describe the computational experience in the development of various delay and cancellation estimation models at the national, regional and airport levels.

A. Grid-Based WITI

WITI indicates how "bad" the weather was based on the number of aircraft affected. It is assumed that traffic and weather information at a given time can be reduced into two two-dimensional grids with the same number of rows and columns. The computation of WITI consists of: 1) assigning a value of one to every grid cell $W_{i,j}$ of the weather grid W where severe weather is indicated and a value of zero elsewhere, 2) counting the number of aircraft in every grid cell $T_{i,j}$, and 3) computing, $X(k)$, the WITI at an instant of time k (typically at one-minute intervals) as follows,

$$X(k) = \sum_{j=1}^{m}\sum_{i=1}^{n} T_{i,j}(k)W_{i,j}(k) \qquad (1)$$

where n is the number of rows and m is the number of columns in the weather grid. The daily national WITI value, X, is given by the summation

$$\boldsymbol{X} = \sum_{\boldsymbol{k}=1}^{1440} \boldsymbol{X}(\boldsymbol{k}), \qquad (2)$$

The en route airspace in the continental United States is divided into 20 geographical areas allocated to individual Air Route Traffic Control Centers (ARTCCs). Given the Center boundary one may calculate the WITI counts within that Center, much the same way as described in equation (1). Let B_p be the closed boundary for Center p and S_p a set of all two dimensional grid cell pair (i, j) inside B_p. Then, the WITI counts for Center p at time instant k can be calculated as

$$\boldsymbol{X}_p(\boldsymbol{k}) = \sum_{(i,j)\in S_p} \boldsymbol{T}_{i,j}(\boldsymbol{k})\boldsymbol{W}_{i,j}(\boldsymbol{k}). \qquad (3)$$

The daily WITI value for Center p, X_p, is given by the summation

$$\boldsymbol{X}_p = \sum_{\boldsymbol{k}=1}^{1440} \boldsymbol{X}_p(\boldsymbol{k}), \qquad (4)$$

B. NAS Weather Index

The WITI metric described in [5], also known as the NAS Weather Index (NWX) when applied to the entire NAS as opposed to individual airports or regions, consists of three components:

- En-route WITI (E-WITI), representing convective weather impact on major flows between city pairs in a linear fashion (amount of convective weather that a flow crosses multiplied by hourly frequency of traffic on this flow);
- Terminal WITI (T-WITI), representing weather impact on major airports, again in a linear fashion (percent capacity degradation multiplied by the number of operations at the airport for the given hour);
- Airport Queuing Delay (Q-Delay), representing surface and terminal-airspace weather impact on major airports in a non-linear fashion (demand related to airport's arrival/departure capacity possibly reduced by weather).

Individual NWX component weights can be obtained via linear regression vs. delay data for a multi-year time period or other methods can be applied, as discussed further in this paper.

IV. ALGORITHMS FOR MODEL DEVELOPMENT

Regression analysis and other data mining techniques are used to model and estimate the various relations. Given X and X_p, two different linear regression models for the national delay, δ, can be developed as,

$$\delta = \alpha \boldsymbol{X} + \beta \qquad (5.a)$$

$$\delta = \sum_{\boldsymbol{p}=1}^{20} \alpha_p \boldsymbol{X}_p + \beta \qquad (5.b)$$

The NAS Weather Index can be replaced as the WITI variable in equations 5.a and 5.b. Subsequently, the two regression models using national WITI, Center/Airport WITIs are referred to as Linear Regression (LR) and Multiple Linear Regression (MLR).

Ref. 6 suggested that the behavior of the NAS is highly nonlinear, and days with higher delays may behave differently from those with lower delays. It should be noted weather reduces airspace capacity by reducing the available resources at airports and in the airspace. In this aspect, the domestic airspace can be viewed as a queuing network, and as the demand for resources reach operational capacity, delay increases exponentially. As an example of the nonlinear nature of delay, for the Atlanta airport, Figure 2 shows a steep increase in delay per operation when the annual number of operations exceed 0.9 million, thus reaching the limit of operational capability. The delay was reduced significantly in 2007 with the addition of a new runway, which increased the operational capacity.

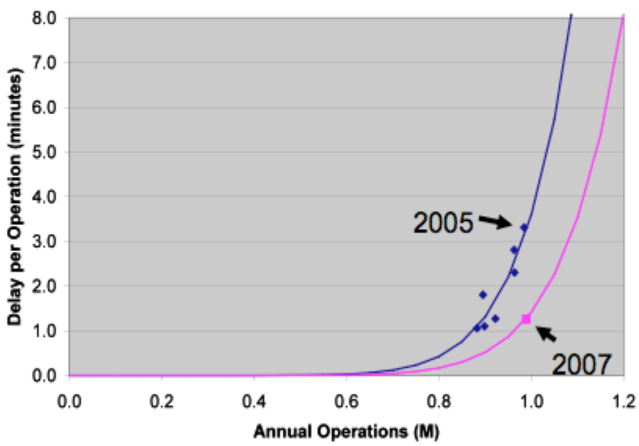

Figure 2. Nonlinear nature of delay vs. number of operations

To capture the nonlinear nature of delay with demand, and to determine if the modeling errors can be reduced further, nonlinear models using piece-wise linear approximation were developed in [8]. A more general approach based on neural networks is examined in the next section.

A. Neural Networks

A satisfactory general-purpose modeling approach allows for modeling a large class of input/output relations, is resistant to noise and missing data, and permits generalization. Generalization of a model is the ability to represent situations not covered during the development phase of the model. Multilayered Neural Networks (NN) have many of these qualities and are used for developing nonlinear prediction models for a wide-range of applications. A major advantage provided by the neural network structure is its use of fairly simple algorithms to learn nonlinear mappings between input/output relations.

A feed-forward neural network consists of input, hidden and output layers and provides a general framework for representing non-linear functional mapping between a set of input variables and a set of output variables. For modeling the NAS performance metrics, the input variables are the 20 Center WITIs and the output variable, also called the target, is the performance metric of interest.

Figure 3 shows a feed-forward neural network with *m* input neurons, a hidden layer with *l* neurons and a single output neuron. The output from each layer is connected to the next layer by modifiable weights represented by links between the layers. The weighted outputs from one layer going through an activation function (a non-linear input-output relation) forms the input to the neuron in the next layer. Nonlinear sigmoid functions are used as activation functions for the hidden neurons in this paper. A bias unit is connected to all neurons except the neurons in the input layer. The weights multiplying the output of the neurons of one layer to the neurons of the next layer are adjusted during the training phase of the neural network development. There are a total of m(*l*+1) weights for the neural network in Figure 3. The back-propagation algorithm based on minimizing the output error using a gradient descent method is used for training neural networks.

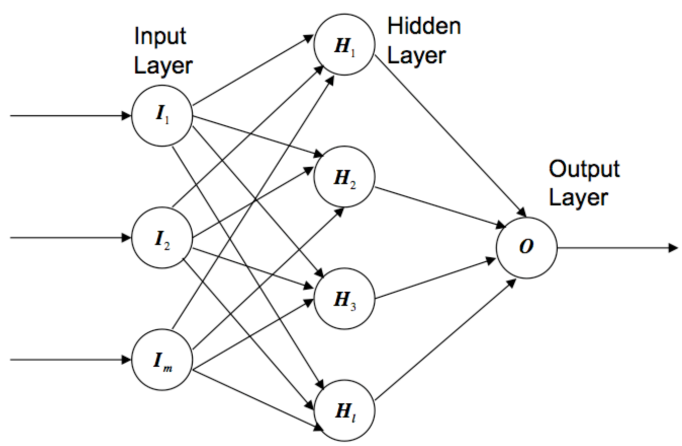

Figure 3. Feed-forward Neural Network.

B. Training and Validation

Supervised learning is used to train the neural network. The training data include both the inputs and the desired outputs. The training procedure starts with a set of initial values for all biases and weights. The entire set of inputs, WITI values of the 20 Centers and the corresponding NAS performance metric on each day in 2004 and 2005, is presented to the NN. The sum of the square of the error (SSE) between the NN network output and the actual observation is computed and the weights are updated using a gradient procedure [14]. At the next epoch, the training is repeated with the new set of weights. The procedure is repeated until the error reaches an acceptable lower bound. The modeling error (SSE), the typical objective function for the training, is reduced as the number of iteration increases. However, minimizing training error can lead to over-fitting and poor generalization if the number of training cases is small relative to the complexity of network.

For the NN to have satisfactory generalization properties, the training should be sufficiently large and statistically representative. The models should not be under-fitted (it could well generalize the test set but poorly approximate the training data) or over-fitted (could approximate the training data well, but poorly generalize the test set). Under-fitting results in models that are too simple or inadequately trained models that have not fully learned the range of input signals. Over-fitting results in models that are too complex and may be trying fit the noise in the signals. To avoid under-fitting and over-fitting, the best number of training data, number of epochs, architecture of the neural network and correct final training state must be determined.

Given a fixed amount of training data, there are several approaches to avoid over fitting, and hence produce satisfactory generalization. One method is Bayesian Regularization (BR), which in addition to minimizing the training error adds a penalty for the complexity of the neural network [15]. A detailed discussion of the use of Bayesian Regularization can be found in [16]. Other approaches for generalization improvement are "Early Stopping" (ES) [17], Principal Component Analysis (PCA) [18] and Stepwise Regression [11].

Neural Network models are data driven and therefore resist analytical or theoretical validation. The models are constructed from an initial random state to a trained state using the training data sets and must be tested or validated using a different data set. In cross-validation, a series of NN models are constructed, each time by dropping a different part of the data from the training set and applying the resulting NN model to predict the target (delays or cancellations). The merged series of predictions for dropped or test data are checked for accuracy against the observation. In one version of the cross-validation approach, called group cross-validation approach, data are divided into N groups. A total of N models are then constructed each using N-1 data groups for model training, and the Nth group for testing. Normally, N can be chosen as 3, 5, and 10. Fivefold cross-validation is used in this paper. The result is the average of the five test results.

A number of methods are available to estimate forecast errors. The two traditional estimates Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are used in this study. MAE and RMSE are measured in the same unit as the original data. MAE is usually similar in magnitude to but slightly smaller than RMSE.

V. NATIONAL MODELS

Several delay models using the WITI concept have been developed in the literature at the national level [1-8]. All but one of these models is linear and use either ASPM or OPSNET data to model a single NAS performance metric. This section describes results on the performance of the linear regression and neural network methods to estimate OPSNET delay,

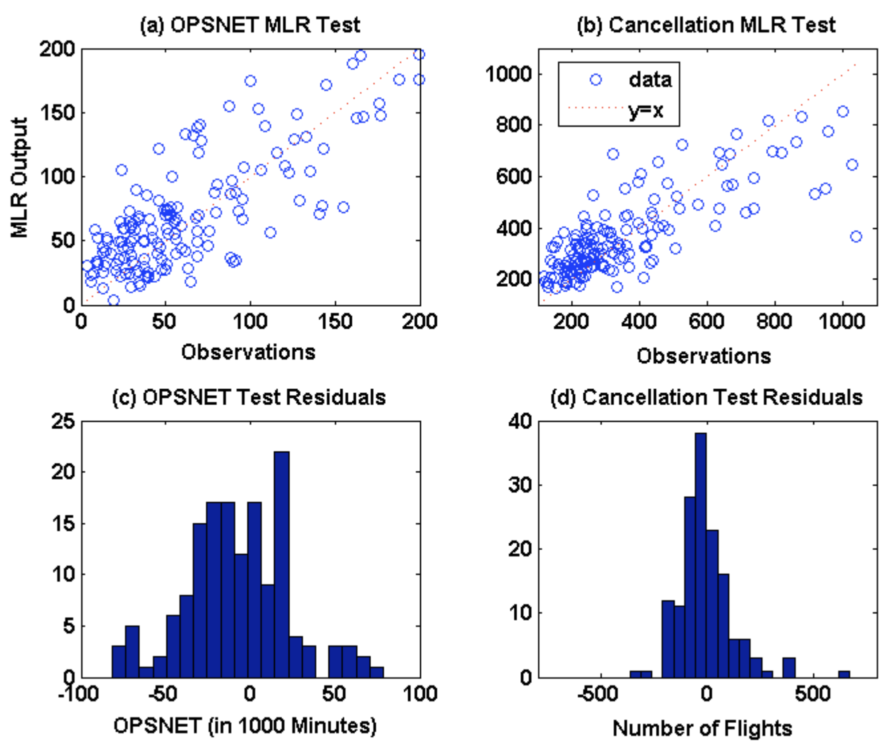

Figure 4 (a) and (b), Scatter plots showing actual and observed metrics, (c) and (d) Distributions for the estimated residuals.

ASPM schedule delay and flight cancellations using 411 days of traffic, weather and metrics during 2004-2006, referred to as the NASA-dataset. The models were developed using weather and delay during 2004 and 2005 and verified using the data for 2006.

Figure 4(a) is a scatter plot of the multiple linear regression model delay estimate, in 1000 minutes, on the y-axis and the observed values of the delay on the x-axis during 2006. Figure 4(b) shows a similar plot for the number of flight cancellations during 2006. The distribution of the error residuals estimates for the aggregate delay and cancellation is shown in Figure 4(c) and (d), respectively. The performance of modeling different metrics is shown in Tables 4.1 – 4.3. The neural networks were designed by reducing their complexity using four different techniques, namely, Bayesian Regularization, Early Stopping, Principal Component Analysis and Stepwise Regression. The performance of the different techniques was similar and results based on Bayesian Regression are reported in this paper. The neural network models were validated using five-fold cross-validation (5C) as described in the previous section.

Table 4.1 Performance of OPSNET national delay models

Type of Model	*r*	RMSE (min)	MAE (min)
LR	.71	32,700	26,600
MLR	.77	31,200	24,500
Neural Network	.80	30,000	23,300
Neural Network (5C)	.80	29,100	22,000

Table 4.2 Performance of ASPM schedule delay models

Type of Model	*r*	RMSE (min)	MAE (min)
LR	.76	97,600	72,900
MLR	.75	99,200	74,300
Neural Network	.80	95,800	74,300
Neural Network (5C)	.79	96,100	73,000

Table 4.3 Performance of ASPM flight cancellation models

Type of Model	*r*	RMSE (flights)	MAE (flights)
LR (5.a)	.73	146	106
MLR (5.b)	.77	131	94
Neural Network	.79	131	93
Neural Network (5C)	.79	139	97

The correlation coefficients and errors for the different metrics in Table 4.1–4.3 indicate that WITI models adequately explain the effect of weather on different types of delays and flight cancellations. The neural networks generally produce models with a higher correlation coefficient and slightly lower modeling errors. The cross-validation results indicate that the resulting neural networks have good modeling and generalization capability for the dataset under study.

A. Performance Metric Estimation with NWX

The behavior of the national delay model was further examined using NWX as the WITI variable and using a larger dataset, referred to as the FAA-dataset, consisting of 1293 days covering the period January 1, 2005 to August 9, 2008. As before, the models were developed using the data for the period 2005–2007 and tested against observations during 2008. The estimation of national delay using NWX is equal to the sum of delays at 34 major airports. The MLR method has 34 regression variables and the neural network has 34 inputs, the daily NWX values at each airport. Table 5 provides a summary of the performance of the models for this dataset.

Table 5. Performance of national delay models

Type of Model	Correlation	
	Training	Testing
LR	.76	.71
MLR	.78	.72
Neural Network	.80	.73

Table 6. Seasonal Performance of national delay models

Type of Model	Summer Correlation		Winter correlation	
	Training	Test	Training	Test
MLR	.85	.75	.74	.71
Neural Network	.85	.76	.75	.72

Table 7. Correlation between performance metrics and WITI

	Air-port Delay	OPS-NET Delay	Flight -plan Delay	Sche-dule Delay	Flight Cancel-lation
E-WITI (LR)	.77	.70	.73	.76	.68
Grid WITI (LR)	.76	.72	.76	.76	.71
NWX (LR)	.84	.80	.84	.84	.74
E-WITI (MLR)	.80	.75	.78	.80	.78
Grid WITI (MLR)	.83	.80	.82	.83	.80
NWX (MLR)	.88	.84	.88	.88	.82

A comparison of Table 4.1 and Table 5 indicates that the performance of the models is similar. WITI is a robust measure, and different definitions of WITI make only slight changes to the performance of the model.

Another question of interest is the behavior of the delay models during different seasons. The data for the period 2005-2008 were divided into two parts consisting of days in April-September (summer model) and days in October-March (winter model). The calculations in Table 6 were performed for these two subsets. Again, the models were developed using the data for the period 2005–2007 and tested against observations during 2008. Lower correlation between WITI and Delay metrics during the winter season may be due to the higher number of cancellations that occur on days with heavy snow, very low ceilings/visibility etc, that are more typical for this season. Delays on such high-impact days are much lower vis-à-vis WITI.

Next, the development of the national delay model explores the relationship between the various definitions of WITI and their influence on the estimation of aggregate delay and cancellation metrics. To study these relations, five different performance metrics, total airport delay, OPSNET delay, flight plan-based ASPM delay, schedule-based ASPM delay and flight cancellations, were estimated using E-WITI, grid-based WITI and NWX. The models were developed using 278 days common to the NASA-dataset and FAA-dataset. Table 7 shows the correlation coefficient between the models for different performance metrics as a function of different definitions of WITI.

For the five different performance metrics in Table 7, using the LR approach, models using NWX perform slightly better than models using either grid-based WITI or E-WITI. However, this small difference between different definitions of WITI is significantly reduced using either MLR or neural network approach.

For the five different performance metrics in Table 7, using the LR approach, models using NWX perform slightly better than models using either grid-based WITI or E-WITI. However, this small difference between different definitions of WITI is significantly reduced using either MLR or neural network approach.

The key results in the development of delay and cancellation models at the national level are: (a) different definitions of WITI (grid-based WITI, E-WITI and NWX) provide similar level of accuracy, (b) the correlation during April-September (~.85) is higher compared to October-March (~.75), (c) generally, correlation (γ) between model predicted delays and actual observed delays is higher for MLR than LR during the training (modeling) phase, and (d) neural networks perform slightly better both during training and testing, and during different seasons compared to regression methods.

VI. AIRPORT DELAY MODELS

This section describes results on the estimation of delays at 34 major U.S. airports. Figure 5 shows the 20 U.S. Air Route Traffic Control Centers (ARTCC) and the 34 major airports that, together with Honolulu (HNL), comprise the OEP-35 list. The NWX was computed for each of these airports during the period 2005–2008.

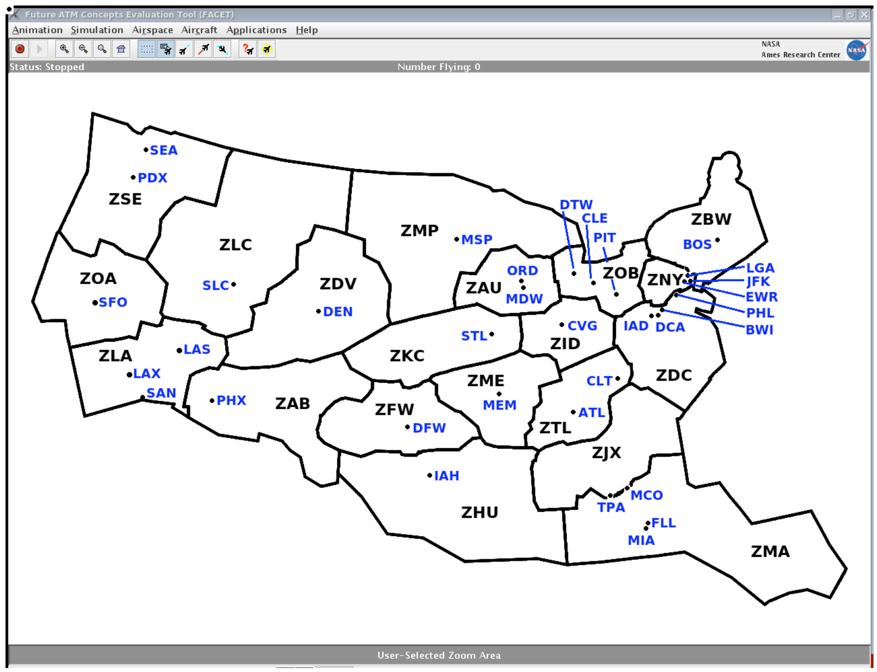

Figure 5. Twenty Centers and location of 34 major airports

The LR, MLR and neural network methods were applied to model delays using the NWX as the WITI variable. The behavior of the LR and MLR models using the entire 2005–2008 data for training is shown in Table 7. γ_{LR} and γ_{MLR} represent the correlation between the estimated delay and the observed delay at an airport using values of NWX at the airport versus using the values of NWX at all the 34 airports.

$$\delta_p = \alpha X_p + \beta \qquad (6.a)$$

$$\delta_p = \sum_{p=1}^{34} \alpha_p X_p + \beta \qquad (6.b)$$

Table 8. Training behavior of LR and MLR airport delay models. The ten airports here are a representative cross-section of the OEP-35 airport group.

Airport	γ_{LR}	γ_{MLR}	γ_{MLR}/γ_{LR} - 1
ORD	0.743	0.803	0.08
ATL	0.752	0.777	0.03
EWR	0.640	0.725	0.13
PHL	0.764	0.805	0.06
DFW	0.577	0.646	0.12
JFK	0.618	0.670	0.08
LGA	0.685	0.723	0.06
LAX	0.195	0.496	1.54
IAH	0.684	0.725	0.06
DEN	0.550	0.664	0.21

The ratio γ_{MLR}/γ_{LR} - 1 in Table 8 is an indicator of the influence of NWX in the neighboring airports on the delay at an airport. As expected, γ_{LR} is less than or equal to γ_{MLR}. Four observations can be made about the airport delay estimates:

(a) For the airports in Table 8, NWX provides good estimates at ORD (.74), ATL (.75), EWR (.64), PHL (.76), JFK (.61), LGA (.69) and IAH (.68);

(b) The delay at ten airports in the Eastern U.S. (ORD, ATL, PHL, JFK, LGA, BOS, DCA, DTW, IAH) are not influenced very highly by the WITI (weather) in the neighboring airports as indicated by the very low values of γ_{MLR}/γ_{LR} - 1, ranging from .0.03 to 0.09;

(c) The delay estimates at certain airports (EWR, DFW, DEN) are moderately influenced by the WITI (weather) in the neighboring airports as indicated by the low values of γ_{MLR}/γ_{LR} - 1; and

(d) The delay estimate using MLR is much higher than the estimate using LR at LAX.

Lower correlation for New York and Philadelphia airports may be due to airspace complexities in that particular area that impact traffic on both good-weather days and weather-impacted days; these complexities are not currently reflected in the WITI/NWX models.

The FAA-dataset has all the days during 2005-2007 and from January 1 to August 2008 during 2008. The delay models using MLR and neural networks were trained using the period 2005-2006 and tested against observations during 2007. Again, to check the variation in the performance of the models with training and test data, the calculations were repeated using the period 2005-2007 for training and delay observations during 2008 for testing. The performance of the models for the different training and test scenarios is shown in Tables 9 and 10.

Table 9. Behavior of airport delay models (Training 2005-2006; Testing 2007)

Airport	Training γ_{MLR}	Testing γ_{MLR}	Training γ_{NN}	Testing γ_{NN}
ORD	.82	.76	.83	.77
ATL	.79	.80	.81	.80
EWR	.78	.64	.80	.68
PHL	.83	.76	.85	.78
DFW	.67	.59	.68	.56
JFK	.69	.60	.71	.63
LGA	.77	.65	.79	.70
LAX	.54	.44	.58	.45
IAH	.78	.60	.80	.60
DEN	.64	.64	.64	.63

Table 10. Behavior of airport delay models (Training 2005-2007; Testing 2008)

Airport	Training γ_{MLR}	Testing γ_{MLR}	Training γ_{NN}	Testing γ_{NN}
ORD	.80	.79	.83	.80
ATL	.79	.72	.80	.72
EWR	.74	.64	.76	.68
PHL	.81	.78	.83	.80
DFW	.65	.60	.69	.63
JFK	.67	.67	.72	.68
LGA	.74	.64	.77	.67
LAX	.54	.34	.55	.35
IAH	.73	.72	.74	.72
DEN	.66	.68	.67	.68

The following observations can be made from the results in the two Tables:

(a) There is no significant difference in the performance of the MLR and neural network models using different non-overlapping periods for testing and training

(b) Neural networks consistently outperform the MLR delay models by a small, but significant amount, both during the testing and the training phase.

VII. CONCLUSIONS

This paper shows that WITI is a good proxy for modeling the impact of weather in the NAS. It answers several important questions on the use of WITI to model delays such as the effect of using different OPSNET versus ASPM database, can some metrics be estimated better than others, and what is the effect of different definitions of WITI. It shows results on the modeling of several NAS performance metrics using data drawn from two operational databases. The paper compares the performance of traditional linear regression models with several neural network models in the estimation of key airspace metrics such as total aggregate delay, arrival delay, and airborne delay and flight cancellations. The performance metrics are predicted at the national, regional and airport levels. The results are based on using the traffic, weather and delay data for the period 2005–2008. Some of the conclusions based on the results of the study are:

(a) The expected number of aircraft impacted by weather is a satisfactory proxy for delay at various levels,

(b) Different delay models are preferable for different seasons, and the delay estimation accuracy is higher in the convective weather season (April-September) vs. the non-convective season (October-March),

(c) Delay estimation accuracy at all levels and for different metrics is about the same,

(d) Models resulting from using either database are complementary and provide the same level of accuracy,

(e) Neural network delay models perform slightly better in that they have a higher correlation between model output and airspace metric then linear regression methods both during development and validation, and during different seasons.

REFERENCES

[1] Callaham, M., DeArmon, J., Cooper, A., Goodfriend, J., Moch-Mooney, D., and Solomos, G., "Assessing NAS Performance: Normalizing for the Effects of Weather," 4th USA/Europe ATM R&D Seminar, Santa Fe, NM, 2001.

[2] Chatterji, G. B., and Sridhar, B., "National Airspace System Delay Estimation Using Weather Weighted Traffic Counts," AIAA Guidance, Navigation and Control Conference, San Francisco, CA, 2004.

[3] Post, J., Bonn, J., Bennett, M., Howell, D., and Knorr, D., "The Use of Flight Track and Convective Weather Densities for National Airspace System Efficiency Analysis," 21st Avionics Systems Conference, Irvine, CA, 2002.

[4] Lamon, K., "Aviation System Performance and Convective Weather," The MITRE Corporation, 2005.

[5] Klein, A., R. Jehlen, and D. Liang, "Weather Index with Queuing Component for national airspace system Performance Assessment," USA/Europe Air Traffic Management R&D Seminar, Barcelona, Spain, July 2007.

[6] Hansen, M., and Xiong, J., "Weather Normalization for evaluating National Airspace System Performance," USA/Europe ATM R&D Seminar, Barcelona, Spain, 2007.

[7] Wieland, F., "Estimating the Capacity of the National Airspace System," 24th Digital Avionics Systems Conference, Washington, DC, November 1-3, 2005.

[8] Sridhar, B., and Swei, S., "Computation of Aggregate Delay Using Center-based Weather Impacted Traffic Index," AIAA Aviation Technology, Integration and Operation Conference (ATIO), Belfast, Northern Ireland, September 2007.

[9] FAA OPSNET and ASPM data are available from http://aspm.faa.gov/. (Note: The ASPM cancellation data of Nov. and Dec. 2006 are missing from the database.)

[10] OAG Travel Solutions, http://www.oag.com/, June 2, 2008.

[11] Draper, N. and Smith, H., "Applied Regression Analysis," 3rd ed. New York: John Wiley & Sons, 1998.

[12] Klein, A., L. Cook, and B. Wood, "Airspace Availability Estimation for Traffic Flow Management Using the Scanning Method," 26th Digital Avionics Systems Conference (DASC), St. Paul, MN, 2008

[13] Lawrence, J. and Fredrickson, J., "BrainMaker User's Guides and Reference Manual," 7th edition, California Scientist Software, Nevada City, CA, 1998.

[14] Duda, R.O., Hart, P.E., and Stork, D.G., "Pattern Classification," John Wiley and Sons, Inc., New York, 2001.

[15] MacKay, D.J.C., "Bayesian interpolation," Neural Computation, Vol. 4, No. 3, pp. 415-447, 1992.

[16] Foresee, F.D. and Hagan M.T., "Gauss-Newton approximation to Bayesian regularization," Proceedings of the 1997 International Joint Conference on Neural Networks, pp. 1930-1935, 1997

[17] Nelson, M.C. and Illingworth, W.T., "A Practical Guide to Neural Nets," Reading, MA: Addison-Wesley, 1991.

[18] Fukunaga, K., "Introduction to Statistical Pattern Recognition," Academic Press, 1990.

AUTHORS' BIOGRAPHIES

Dr. Banavar Sridhar worked at Systems Control, Inc., Palo Alto, Ca and Lockheed Palo Alto Research Center before joining NASA Ames Research Center in 1986. He was recently appointed as NASA Senior Scientist for Air Transportation Systems. His research interests are in the application of modeling and optimization techniques to aerospace systems. Dr. Sridhar received the 2004 IEEE Control System Technology Award for his contributions to the development of modeling and simulation techniques. He led the development of traffic flow management software, Future ATM Concepts Evaluation Tool (FACET), which received the NASA Software of the Year Award in 2006. He is a Fellow of the IEEE and the AIAA.

Dr. Yao Wang is a Computer Scientist at NASA Ames Research Center, specializing in statistical data analysis and data mining. He has worked for the past 8 years on Aviation related researches and applications. He earned his Ph.D. degree in Physics from University California of Santa Barbara in 1988.

Dr. Alexander Klein has over 20 years experience in creating airport and airspace simulation models, airspace analysis tools, airline decision support tools, weather translation models, as well as in NAS performance analysis, computer graphics and operations research. His current customers include the FAA, NASA, MITRE, and the National Weather Service. Past customers include leading civil aviation authorities, airlines, airports and aerospace companies in the US and in over 20 countries.

Richard Jehlen is Director - Integration for System Operations Services in the FAA's Air Traffic Organization. His office serves as the principal Air Traffic element responsible for the NAS and ICAO operational concepts. It is also responsible for Air Traffic's validation of the concepts, international harmonization efforts for the future and ultimately the efficient integration of new systems. Over 22 years in ATC, his responsibilities have included Airspace & Procedures, Military Liaison, Automation and direct operational oversight.