Lessons Learned in the Application of Machine Learning Techniques to Air Traffic Management

Banavar Sridhar *, Gano Chatterji† and Antony Evans‡
NASA Ames Research Center, Moffett Field, CA, USA

There is an increasing interest in applying methods based on Machine Learning Techniques (MLT) to problems in Air Traffic Management(ATM). The current interest is based on developments in Cloud Computing, the availability of open software and the success of MLT in automation, consumer behavior and finance involving large databases. This paper reviews the current-state-of-the art in applying MLT to aviation operations, its promises and challenges. Historically aviation operations have been analyzed using physics-based models and provide information for making operational decisions. Aviation operations involving many decision makers, multiple objectives, poor or unavailable physics-based models and a rich historical database are prime candidates for analysis using data-driven methods. The promises and challenges in applying MLT to ATM is traced through three examples based on the authors' experience, each separated by a decade, to show the influence of data and feature selection in the successful application of MLT to ATM. As always, the best approach depends on the task, the physical understanding of the problem and the quality and quantity of the available data.

I. Introduction

Modeling and analysis of problems in aviation operations has a rich history dating back to the early nineteen 1970s[1]. Federal Aviation Administration (FAA)'s Automated En-Route Air Traffic Control (AERA) concept was developed to provide prediction, planning, optimization and decision-making techniques to increase fuel efficiency, increase controller productivity, and to reduce system errors [2]. It provides an example of wide use of modeling and analysis techniques in aviation operations. The task or the question to be answered by the analysis determines the approach and the methodology used to solve the problem.Traditionally the analysis methods are based on concepts developed in Guidance, Navigation and Control, Optimization Theory, Statistics, Parameter Estimation and Numerical Analysis. Further, the solution depends on the available models to represent the problem, the ability to capture the objective of the task in a cost function and optimization or numerical approach to present a solution to help in decision making. Many problems in aviation operations like Conflict Detection and Resolution (CDR), scheduling of departures and arrivals in airports, Traffic Flow Management (TFM) and modeling of weather uncertainties have been addressed using traditional methods. Recent developments in Cloud Computing, availability of open software and the success of Machine Learning Techniques (MLT) in automation, consumer behavior and finance involving large databases has led to an increase in the application of MLT to analyze problems in Aviation Operations (AO). The progress in the application of MLT to Air Traffic Management (ATM) is traced through three examples based on the authors' experience, each separated by a decade, to show the influence of data and feature selection in the successful application of MLT to ATM.

The paper is organized as follows: Section II, Simulation and Analysis describes the steps and resources in the modeling, simulation and optimization techniques used in aviation operations. Section III, Machine Learning Techniques provides a description of the common MLT that have been applied to aviation problems. Section IV, Data Sources provides the types of database available for the solution of problems in aviation operations. Section V, Application of MLT provides three examples, (1) air traffic complexity, (2) estimation of delay and cancellations in the US National Airspace and (3) reroute advisories, to illustrate the role of data and selection of features in the application of MLT to different problems in ATM. Section VI, Conclusions provides concluding remarks and future beneficial roles for MLT to address problems in aviation operations.

*Principal Scientist, USRA@NASA Ames Research Center, Moffett Field, CA 94035, USA , Fellow AIAA
†Senior Scientist and Lead, Crown Consulting Inc., NASA Ames Research Center, Moffett Field, CA 94035, USA, Associate Fellow AIAA
‡Scientist, Crown Consulting Inc., NASA Ames Research Center, Moffett Field, CA 94035, USA.

II. Simulation and Analysis

This section describes some of the components of a simulation and analysis system to support decision-making by the stakeholders of aviation operations, namely airlines, traffic service providers and regulatory bodies to provide a safe, efficient and affordable air transportation system to the public. Many of the important problems in aviation operations can be formulated using a dynamic system described by a set of equations

$$\frac{dx}{dt} = f(x, u, w, \theta)$$

$$y = g(x, u, w, \theta)$$

where x is the state of the system, u is the control, y is output of the system, w disturbances or uncertainty affecting the system and θ are a set of known or unknown parameters. An operation on the system may be to make the system output follow the desired output $y_d(t)$ during the time interval t_0 to t_f. This might be achieved by minimizing the function

$$\underset{u}{Min} \int_{t_0}^{t_f} (y_d - y)^2 + u^2 dt$$

w.r.t the control variable $u(t)$. In the discrete formulation, time is divided into several intervals resulting in a multi-stage optimization problem

$$\underset{u}{Min} \sum_{i} (y_d(i) - y(i))^2 + u(i)^2$$

and the optimal cost $J^*(i)$ satisfies Bellman's equation (1) where $c(i, u, j)$ is the transition cost (reward) from stage i to j and $J^*(j)$ is referred to as the cost-to-go for the remaining stages.

$$J^*(i) = \underset{u}{Min} \quad E[c(i, u, j) + J^*(j)|i, u] \quad for \quad all \quad i \tag{1}$$

This framework can be used to solve the task as an optimization problem or as a classification problem.

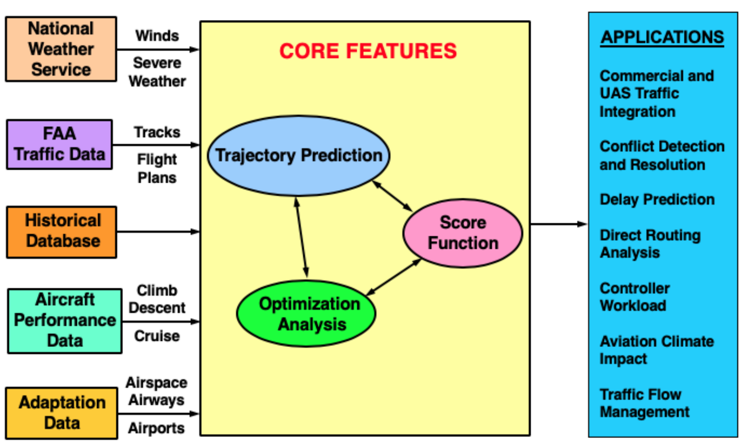

Fig. 1 Components of Simulation and Analysis

The solution to optimization or classification problems in ATM requires simulation of air traffic scenarios. Fig 1 shows some of the components in a simulation software for exploration, development, and evaluation of advanced Air Traffic Management (ATM) concepts [3]. The simulation models system-wide airspace operations over the contiguous United States. Airspace models (e.g., Center/Sector boundaries, airways, locations of navigation aids and airports) are available from the FAA database. Weather models (winds, temperature, severe weather cells, etc.) are available from the National Weather Service in the US. Aircraft trajectories are modeled using 3 Degree-of-Freedom equations of motion. The aircraft can be flown along their routes as they climb, cruise, and descend according to their individual aircraft-type performance models. These 4D trajectories provide the engine that drives various ATM applications. More detail about the various inputs to the simulation and the database providing the information is discussed in the section on Database.

The level of detail of the various components and approach used in simulation and analysis of ATM systems is dependent on a clear statement of the task or goal. The task may vary from providing conflict detection and resolution support to air traffic controllers [4], to providing policy options to reduce the impact of aviation on climate [5]. The simulation scenarios vary in time ranging from a few seconds to decades, from controlling small regions of airspace to global air traffic management and involving expertise in aeronautics, atmospheric sciences, human factors, optimization and data sciences.

A. Physics-based Modeling and Data-driven Modeling

Physics-based models use knowledge about the relationship between physical quantities based on prior science and experience. On the contrary, data-driven models treat input-output relations as a black-box and try to establish the relationship between the variables based solely on data. A combination of the two methods can be used depending on the task. Table 1 summarizes some of the similarities and differences between physics-based and data-driven approaches to modeling complex systems. MLT provides a new set of tools to model the complex problems in aviation operations.

Table 1 Comparison of Physics-based and Data-driven Models

Property	Physics-based Models	Data-driven Models
Model	Linear, Non-Linear, Dynamic, Static, Queueing	Black-Box
Interpretation	Easy to explain results in terms of physical quantities	Can be hard to interpret and gain trust in the system
Model-Building	Expensive and requires lot of application expertise	Availability of quantity and quality of data
Suitability	Availability of well-defined physical models	Ideal for building causal relationship between inputs and outputs when good physics-based models are non-existent or expensive to build
Feature Selection	Defined by the model and various methods to reduce dimensions (Aggregation, time and space separation)	Major issue to reduce the dimension in complex problems
Size	Various methods to determine minimal order unbiased minimal variance models	Efforts to balance over-fitting and under-fitting by cross-validation, regularization and other methods

III. Machine Learning Techniques

In this study, we use three data mining methods: ensemble bagging decision trees (BDT), neural network classifier and support vector machine (SVM) learning algorithms. These models were selected because decision surfaces modeled by these approaches are very different from each other. As we do not have direct knowledge of the shape of decision surface in the case of ATM applications, it would be good to use different data mining methods to examine if the actual surface is modeled more accurately by one of these methods.

A. Support Vector Machine

Linear Classifiers using Linear Discriminant Functions (LDF) are very attractive due to their computational simplicity and as a starting point for classification. LDF is computed by minimizing the error in the classification of training samples or observations. It should be noted that a small training error does not guarantee a good performance in classifying a general sample. Many descent procedures are used to reduce the classification error and find the hyperplane describing the LDF [6]. A support vector machine [7] constructs a hyperplane or set of hyperplanes in a high-dimensional space, which can be used for classification and regression. Its robust performance with respect to limited, sparse and noisy data is making it widely used in many applications from protein function, face recognition and text categorization. The SVM model has also been utilized in airport capacity classification prediction. When used for binary classification, the SVM algorithm separates a given set of two-class training data by constructing a multidimensional hyperplane that optimally discriminates between the two clusters. Although SVMs were originally proposed to solve linear classification problems, they can be applied to non-linear decision functions by transforming the inputs using kernel functions. The hyperplane in the high dimensional space corresponds to a non-linear decision boundary in the input space. A widely used kernel is the Gaussian radial basis function (RBF)[8].

B. Ensemble Bagging Decision Tree Classifier

Decision tree learning [9] uses a decision tree as a predictive model that associates input variables with target values. Each internal node corresponds to a condition on an input variable; there are edges from the node to children for each of the possible values of that input variable. Each leaf node has a value of the target variable associated with it. This value is the predicted value given the values of the input variables represented by the path from the root of the tree to the leaf. Algorithms for creating decision trees work top-down by selecting a condition on a variable at each step that best splits the set of items. In this study , the algorithm used a metric called "Gini impurity." Gini impurity is a measure of how often a randomly chosen element from the set would be incorrectly classified if it were randomly classified according to the distribution of classes in the subset. Another commonly used measure is the "information gain" measure. Ensemble methods use multiple machine learning models to obtain better predictive performance than what any of its individual constituent members can produce. Bagging is an ensemble method that uses random re-sampling of a dataset to construct models.

C. Neural Network Classifier

The class of solutions provided by hyperplane decision boundaries, while applicable to solve a large class of problems, is not sufficient to reduce the classification error in some complex applications involving nonlinear boundaries. Multilayered Neural Networks (NN) provide a satisfactory general-purpose modeling approach 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. 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 feedforward neural network [10] consists of nodes in the 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. The output from the nodes of each layer is connected to the nodes of the next layer by modifiable weights represented by links between the layers. The weighted outputs from each node goes through nonlinear sigmoid functions to form the input to the nodes in the next layer. A bias unit is connected to all nodes except the nodes in the input layer. The backpropagation algorithm based on minimizing the output error using a gradient descent method is used for training neural networks.

Figure 2 shows a feed-forward neural network with m input nodes, a hidden layer with l node and a single output node. The weights 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 2.

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 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. At the next epoch, the training is repeated with the new set of weights. The procedure is repeated until the error converges to 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 size of the network. For the NN to have satisfactory generalization properties, the training should be sufficiently large and statistically representative. Under-fitting results in models that are too simple and 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 [11]. A detailed discussion of the use of Bayesian

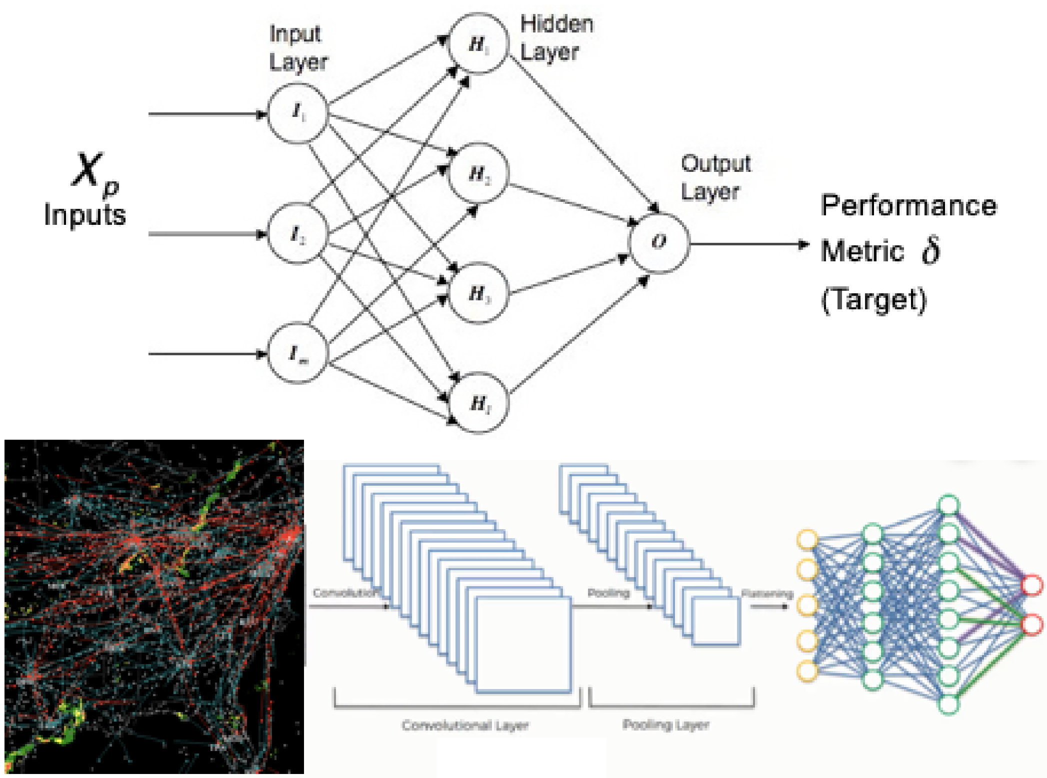

Fig. 2 Feed-forward neural network

Regularization can be found in [12]. Other approaches for generalization improvement are "Early Stopping" (ES)[13], Principal Component Analysis (PCA)[14] and Stepwise Regression [15].

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 output or target. 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. A number of methods are available to estimate forecast errors. The two traditional estimates are Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

IV. Data Sources

A. NAS Performance Metrics and Database

The FAA maintains several databases to measure the demand and efficiency performance in the National Airspace System (NAS). Two of them, Operations Systems Network (OPSNET) and Aviation System Performance Metrics (ASPM), include measures of air traffic control system performance.

1. OPSNET

OPSNET collects data including 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 during a day. The OPSNET delays are further categorized by flight category, class of traffic and cause of delay.

2. ASPM

ASPM covers 77 major airports in the United States. It provides the airport specific data, runway configuration and the local meteorological conditions at each airport. Hourly values of wind speed, visibility, ceiling, Instrument Meteorological Conditions (IMC), scheduled arrivals and departures, airport hourly delays and airport arrival rates (AAR), from ASPM data are used as input variables in several studies. The ASPM delays are measured relative to flight plans 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 worldwide airline schedules database. AAR is a dynamic parameter specifying the number of arrival aircraft that an airport can accept in an hour. Actual hourly airport surface weather observation reports (METAR) including wind, ceiling, visibility, and meteorological condition are used by air traffic controllers and by meteorologists. The processed ASPM data is used in the statistical analysis and also as inputs for generating and validating the machine learning models. The delays are based on flight data from the Traffic Flow Management System (TFMS) 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. OPSNET and APM contain data entered by human operators and are prone to data recording errors. The two databases are overlapping in certain areas and complementary in others. Both databases can be used independently for developing NAS metrics models based on statistical analysis.

B. Bureau of Transportation Statistics (BTS)

The U.S. Department of Transportation requires air carriers to report on domestic operations to and from U.S. airports. Data from these reports are made available by the Bureau of Transportation Statistics (BTS). All Carriers that have more than 0.5 percent of total domestic scheduled-service passenger revenue report on-time data and the causes of delay. The airlines report the causes of delays in five broad categories: (a) air carrier delays: cancellation or delay due to circumstances within the airline's control (e.g. aircraft maintenance or crew problems); (b) extreme weather conditions such as tornado, blizzard or hurricane that delays or prevents the operation of a flight; (c) NAS: delays and cancellations attributable to the national aviation system to manage traffic safely during non-extreme weather conditions, airport operations and heavy traffic volume; (d) late-arriving aircraft: Flight delayed due to aircraft arriving late from a previous flight; (e) security: delays or cancellations due to security related events.

V. Application of MLT

This section reviews the application of MLT to problems in aviation operations. The review is done by providing three applications of MLT to ATM problems: (1) air traffic complexity, (2) estimation of delay and cancellations, and (3) reroute advisories. These applications provide factors to consider for successful application of MLT to ATM problems.

A. Air Traffic Complexity

Current ATM is human-centric and technology for fully autonomous operations do not exist. For increasing automation in a safety-critical system such as ATM, system evolution and the introduction of automation and autonomy should be understandable and acceptable to human operators with graceful degradation during failures or off-nominal conditions [16]. The safety of aircraft in the airspace is maintained by the cooperation between pilots, air traffic controllers and airline dispatchers. A Sector controller ensures that all aircraft are separated by a certain minimum distance in the Sector. The controllers perform their task with a combination of displays and decision aids. The amount of traffic a controller can manage safely in a Sector is expressed in terms of maximum number of aircraft , Monitor Alert Parameter (MAP) of a Sector. Many efforts have been undertaken to replace MAP with a more suitable metric, such as Dynamic Density, for describing the complexities of the traffic in a Sector. Airspace complexity and its determination is even more relevant in future air traffic applications like Urban Air Mobility [17]. Airspace complexity, due to lack of physics-based models, provides a good example for the application of MLT to ATM problems.

Airspace complexity [18] depends on both structural and flow characteristics of the airspace. The structural characteristics are fixed for a Sector/Center and they depend on the spatial and physical attributes of the Sector such as terrain, number of airways, airway crossings and navigation aids. The flow characteristics vary as a function of time and depend on features like number of aircraft, mix of aircraft, weather, separation between aircraft, closing rates, aircraft speeds and flow restrictions. A combination of these structural and flow parameters influences the controller workload.

Chatterji and Sridhar [19] present an early example of the application of a neural network based method for mapping the measures of second-order statistics derived from the air traffic data to the qualitative workload assessment by an air traffic controller. To compute the measures of second-order statistics of the kinematic variables, a notion of the spatial distribution of the aircraft within the airspace was required. This notion was established by using the minimum-spanning-tree method. It was shown that the neural network is trainable with the measures of second-order statistics. The trained neural network is able to predict controller workload with a high degree of certainty.

Air traffic data for aircraft within the Dallas Fort Worth (DFW) Air Route Trafic Control Center (ARTCC) were recorded on August 10, 1998 between 4 and 6 pm using the Center TRACON Automation System (CTAS) [4]. The recorded data was played back using CTAS and an air traffic controller from DFW who controls traffic within Sector 86 airspace rated the workload for the same Sector using the graphical interface. The input workload assessment was recorded with a time stamp available in the prerecorded traffic file. The controller who participated in the workload assessment experiment were of the opinion that the workload should be categorized in three levels: low workload, medium workload and high workload. However the majority of the dataset contained only the low and medium workload ratings.

At each discrete time separated by a 12 second interval, 12 scalar measures were computed for each of the six kinematic variables, the three components of the position and velocity vectors, using the procedure described in [19]. Thus, 72 inputs are used for training the neural network similar to the neural network in Figure 2. The three-layer neural network with 72 nodes in the input layer, 15 nodes in the hidden layer and three nodes in the output layer was trained using the standard gradient-based backpropagation algorithm. The three output nodes are designed to output [0.1, 0.1, 0.91] for low workload assessment, [0.1,0.9, 0.11] for medium workload assessment and [0.9, 0.1, 0.11] for high workload assessment. The controller workload ratings used for training are also provided in the same form.

The neural network correctly predicted workload situation 95% of the time and the medium workload situation 82% of the time. The network falsely predicted the medium workload situation as a low workload situation 18% of the time and the low workload situation as a medium workload situation 5% of the time. The false alarm rate is expected to decrease if more data is used for training the neural network. Some of the characteristics of the method are summarized in Table 2.

Table 2 Air Traffic Controller Workload [19]

Problem	**Air Traffic Controller Workload**
Data	Recorded air traffic data (position, velocity of aircraft) at Fort Worth Center Aug 10, 1998
Method	Gradient-based Back-propagation Neural Network
Feature Selection	Spatial distribution of aircraft represented by a minimum spanning tree
Method of Evaluation	MAE, RMSE and Correlation Coefficient Confusion Matrix
Remarks	NN correctly identified 95% of low-workload cases, 82% medium-workload cases and was unable to identify high-workload cases due to limited high-workload samples

B. National Airspace Performance Metrics

As described earlier, the actions taken by the FAA to maintain safety in the presence of excess traffic demand for airspace, especially during convective weather, results in delays. The availability of a delay database and the importance of improving the effectiveness of Ground Delay Programs (GDP) has resulted in several efforts to understand the connection between weather and delay at the local and national levels [2-8]. Delay models have been developed using linear and non-linear regression models, and performance metrics have originated from using either OPSNET or ASPM. Sridhar et al [20] compare 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. These metrics are predicted at the national, regional and airport levels. The results are based on using the traffic, weather and delay data, from both OPSNET and ASPM, for the period 2004–2008. Given the images of traffic data and weather data, convolutional neural networks with their ability to aggregate data in the convolutional and pooling layers could be a natural choice to model this problem. However, the selection of the Weather Impacted Traffic Index (WITI), a measure of the number of aircraft affected by weather at a given time, performs the processing done by the convolutional layers while decreasing the computational time. 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 (typically at one-minute intervals) as follows,

$$X(k) = \Sigma_{j=1}^{m} \Sigma_{i=1}^{n} T_{ij}(k) W_{ij}(k)$$

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

$$X = \Sigma_{k=1}^{1440} X(k)$$

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

$$X_p(k) = \Sigma_{(i,j) \in = S_p} T_{ij}(k) W_{ij}(k)$$

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

$$X_p = \Sigma_{k=1}^{1440} X(k)$$

Given $X$and X_p, two different linear regression models for the national delay, δ can be developed as,

$$\delta = \alpha X + \beta$$

$$\delta = \Sigma_{p=1}^{20} \alpha_p X_p + \beta$$

The results from Linear Regression (LR) and Multiple Linear Regression (MLR) will be compared with the results from NN later in the paper.

The feedforward neural network used for comparison has 20 Center WITI values in the input layer and a single node in the output layer. 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]. The modeling error (SSE), the typical objective function for the training, is reduced as the number of iterations increases.

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).

A number of methods are available to estimate forecast errors. The two traditional estimates MAE and 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.

This section describes results on the performance of the linear regression and neural network methods to estimate OPSNET delay, 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 developed using weather and delay during 2004 and 2005 were verified using the data for 2006. The performance of modeling different metrics are shown in Tables [3,4 and 5]. Some of the characteristics of the methods are summarized in Table 6.

Table 3 Performance of OPSNET national delay models [20]

Type of Model	r	RMSE (min)	MAE (min)
LR	0.71	32,700	26,000
MLR	0.77	31,200	24,500
Neural Network	0.80	30,000	23,300
Neural Network (5C)	0.80	29,100	22,000

Table 4 Performance of ASPM schedule delay models [20]

Type of Model	r	RMSE (min)	MAE (min)
LR	0.76	97,600	72,900
MLR	0.75	99,200	74,300
Neural Network	0.80	95,800	74,300
Neural Network (5C)	0.79	96,100	73,000

Table 5 Performance of ASPM flight cancellation models [20]

Type of Model	r	RMSE (flights)	MAE (flights)
LR	0.73	146	106
MLR	0.77	131	94
Neural Network	0.79	131	93
Neural Network (5C)	0.79	139	97

Table 6 Performance of ASPM flight cancellation models[20]

Problem	**Flight Delay and Cancellation in US**
Data	FAA OPSNET and ASPM data Convective weather 2005-2008.
Method	MLR and feed-forward NN with several stopping criteria
Feature Selection	Weather Influenced Traffic Index(WITI) at the Center, National and airport level
Method of Evaluation	MAE, RMSE and Correlation Coefficient
Remarks	For all metrics and seasons at all levels NN produced slightly better results the MLR

C. Reroute Advisories

Dynamic Weather Routes (DWR)[21] is a trajectory automation system that continuously and automatically analyzes trajectories of flights en-route to find simple modifications to their current routes that can save significant flying time and which are more likely to be acceptable to the pilot and controller, while avoiding weather, traffic conflicts and airspace Sector congestion. DWR users, including airline flight dispatchers, are alerted when a route change for a flight can potentially save more than a user specified minimum amount of flight time. Flight Dispatchers can visualize proposed reroutes, modify them if necessary, evaluate key parameters and provide the route modification to pilots for further consideration. After reviewing and accepting the proposed reroute, the pilot can verbally make the reroute request to the air traffic controller operating the Sector in which the aircraft is located at that time.

A number of studies have examined elements of route acceptability to Air Traffic Control (ATC). Generally, the factors used to evaluate the operational acceptability of routes are identified through observations and feedback from Subject Matter Experts (SMEs). However, the ability of these factors to predict route operational acceptability may be limited and has not been validated. Air traffic controller decision-making varies significantly under different conditions, and it is difficult, even for SMEs, to explain all the drivers of their decisions. Operational data, however, if mined effectively, has the potential to capture many of these complexities. A predictor of the operational acceptability of reroute requests, trained on historical operational data, could capture the impact of different factors for which features can be calculated, and how these different factors contribute to the overall acceptance or rejection of a reroute request. They also provide the opportunity to identify the dominant drivers of ATC acceptability from the predictor feature set.

The operational trial of DWR concept [22] at American Airlines provided the training and test data needed for developing the predictors of route acceptability in this paper. Ten features made up of characteristics describing reroute usage, congestion along the reroute, reroute deviation and the reroute starting point were used to develop the algorithms. These features cover the majority of those identified in the literature.

The acceptability of DWR is formulated as a two-class classification problem. A number of algorithms were applied to train the model, using the R statistical computing environment: (1) logistic regression, (2) support vector machine using a sigmoid kernel, (3) single decision tree, which uses a decision tree structure to classify data, and two ensemble methods – (4) random forest and (5) Adaptive boosting (AdaBoost) [23]. Ten-fold cross validation was used to estimate performance using each algorithm. A number of measures, accuracy, misclassification error, true positive rate, true negative rate, precision, F-score, area under ROC and average precision, were used to compare the algorithms [24]. The best performing model in terms of F-Score is the random forest (0.815), with AdaBoost the second best (0.766). The random forest also performs best under all other metrics. These results are generally expected as ensemble learning techniques such as a random forest and AdaBoost typically outperform other machine learning algorithms because the group of classifiers trained performs more accurately than any single classifier, utilizing the strengths of the individual group of classifiers while at the same time circumventing the weaknesses of the individual classifier [25]. Some of the characteristics of the method are summarized in Table 7.

VI. Conclusions

This paper reviewed some of the challenges in modeling aviation system operations. Over time the tools for predicting the behavior of a large complex system and make decisions to improve the performance of the system has varied from physics-based models to data-driven models to a combination of physics-based and data-driven models. The various issues arising in the modeling of different tasks in aviation operations are examined by applying the techniques to the problems of controller workload estimation, delay estimation and reroute selection. The examples reveal that neural networks generally performed better than regression methods in estimating delays in the NAS. Insufficient data in high workload situations led to underfitting in workload estimations. Random Forest methods performed well in reroute applications. This suggests that there is no single MLT technique which is most suitable for all applications. The selection of features makes a big difference in approximating the non-linear optimization function involving several thousand variables. Neural networks provide a wide variety of approximations. MLT provides a complimentary set of tools that should be considered in applications and the choice of the appropriate method depends on the task, pre-existing knowledge, expert opinion and available data sources.

Table 7 Reroute Advisories [24]

Problem	**Reroute Advisories**
Data	Trial of DWR concept at American Airlines Accepted and rejected reroute advisories during May-September, 2014
Method	Logistic Regression, SVM, Decision Tree RF and Adaptive Boosting
Feature Selection	10 Feautures based on controller and pilot activity and expert opinion
Method of Evaluation	MAE, RMSE and Correlation Coefficient Confusion Matrix, F1 and ROC Ten-fold cross-validation
Remarks	RF and Adaboost performed best with F1 score 0.815 and 0.766 respectively better on estimating delay on individual links. Performs varies with problem and prediction horizon

Acknowledgments

The author acknowledges the contributions and support of many colleagues at NASA, USRA, ENAC and NTU in the development of this paper.

References

[1] McClamroch, H., and Sridhar, B., "Guest Editorial, Special Issue on Automated Air Traffic Control Systems," *IEEE transactions on intelligent transportation systems*, Vol. 2, No. 2, 2001, pp. 37–38.

[2] Goldmuntz, L., Kefaliotis, J. T., Kleiman, L. A., Rucker, R. A., and Schuchman, L., "Automated en-route air traffic control (AERA) concept," Tech. rep., Federal Aviation Administration,Washington,DC. Office of Systems Engineering, Tech. Report, 1981., 1981.

[3] Bilimoria, K. D., Sridhar, B., Grabbe, S. R., Chatterji, G. B., and Sheth, K. S., "FACET: Future ATM concepts evaluation tool," *Air Traffic Control Quarterly*, Vol. 9, No. 1, 2001, pp. 1–20.

[4] Erzberger, H., Lauderdale, T. A., and Chu, Y.-C., "Automated conflict resolution, arrival management and weather avoidance for ATM," 2010.

[5] Sridhar, B., Chen, N. Y., Ng, H. K., and Morando, A., "Modeling and simulation of the impact of air traffic operations on the environment," *Air Traffic Control Quarterly*, Vol. 20, No. 4, 2012, pp. 285–310.

[6] Duda, R. O., Hart, P. E., and Stork, D. G., "Pattern Classification New York," *NY: Wiley-Interscience*, 2001.

[7] Cortes, C., and Vapnik, V., "Support-vector networks," *Machine learning*, Vol. 20, No. 3, 1995, pp. 273–297.

[8] Hand, D., Mannila, H., and Smyth, P., "Principles of data mining. 2001," *MIT Press. Sections*, Vol. 6, 2001, pp. 2–6.

[9] Quinlan, J. R., "Induction of decision trees," *Machine learning*, Vol. 1, No. 1, 1986, pp. 81–106.

[10] Chauvin, Y., and Rumelhart, D. E., *Backpropagation: theory, architectures, and applications*, Psychology Press, 2013.

[11] Mckay, D., "Information-based objective functions for active data selections," *Neural Computation*, Vol. 4, No. 4, 1992, pp. 590–604.

[12] Forsee, F. D., and Hagan, M., "Gauss-Newton approximation to Bayesian learning," *1997 IEEE international conference on neural networks*, 1997, pp. 1–4.

[13] Nelson, M. M., and Illingworth, W. T., "A practical guide to neural nets," 1991.

[14] Fukunaga, R., *Statistical pattern recognition*, Academic Press., 1990.

[15] Draper, N. R., and Smith, H., *Applied regression analysis*, Vol. 326, John Wiley & Sons, 1998.

[16] Edwards, T. E., and Lee, P. U., "Towards Designing Graceful Degradation into Trajectory Based Operations: A Human-Machine System Integration Approach," *17th AIAA Aviation Technology, Integration, and Operations Conference*, 2017, p. 4487.

[17] Bulusu, V., Sengupta, R., and Liu, Z., "Unmanned Aviation: To Be Free or Not To Be Free? A Complexity Based Approach," *7th International Conference on Research in Air Transportation*, 2016.

[18] Sridhar, B., Sheth, K. S., and Grabbe, S., "Airspace complexity and its application in air traffic management," *2nd USA/Europe Air Traffic Management R&D Seminar*, 1998, pp. 1–6.

[19] Chatterji, G. B., and Sridhar, B., "Neural network based air traffic controller workload prediction," *Proceedings of the 1999 American Control Conference (Cat. No. 99CH36251)*, Vol. 4, IEEE, 1999, pp. 2620–2624.

[20] Sridhar, B., Wang, Y., Klein, A., and Jehlen, R., "Modeling flight delays and cancellations at the national, regional and airport levels in the United States," *8th USA/Europe ATM R&D Seminar, Napa, California (USA)*, 2009.

[21] McNally, D., Sheth, K., Gong, C., Love, J., Lee, C. H., Sahlman, S., and Cheng, J., "Dynamic weather routes: a weather avoidance system for near-term trajectory-based operations," *28th International Congress of the Aeronautical Sciences*, 2012, pp. 23–28.

[22] McNally, D., Sheth, K., Gong, C., Sterenchuk, M., Sahlman, S., Hinton, S., Lee, C. H., and Shih, F.-T., "Dynamic weather routes: two years of operational testing at American Airlines," *Air Traffic Control Quarterly*, Vol. 23, No. 1, 2015, pp. 55–81.

[23] Schapire, R. E., "Explaining adaboost," *Empirical inference*, Springer, 2013, pp. 37–52.

[24] Evans, A. D., Sridhar, B., and McNally, D., "Improving Operational Acceptability of Dynamic Weather Routes Through Analysis of Commonly Use Routings," *16th AIAA Aviation Technology, Integration, and Operations Conference*, 2016, p. 3600.

[25] Kotsiantis, S., and Pintelas, P., "Recent advances in clustering: A brief survey," *WSEAS Transactions on Information Science and Applications*, Vol. 1, No. 1, 2004, pp. 73–81.