

Modeling Weather Impact on Airport Arrival Miles-in-Trail Restrictions

Yao Wang

NASA Ames Research Center

Shon Grabbe

National Aero & Space Administration

ABSTRACT

When the demand for either a region of airspace or an airport approaches or exceeds the available capacity, miles-in-trail (MIT) restrictions are the most frequently issued traffic management initiatives (TMIs) that are used to mitigate these imbalances. Miles-in-trail operations require aircraft in a traffic stream to meet a specific inter-aircraft separation in exchange for maintaining a safe and orderly flow within the stream. This stream of aircraft can be departing an airport, over a common fix, through a sector, on a specific route or arriving at an airport. This study begins by providing a high-level overview of the distribution and causes of arrival MIT restrictions for the top ten airports in the United States. This is followed by an in-depth analysis of the frequency, duration and cause of MIT restrictions impacting the Hartsfield-Jackson Atlanta International Airport (ATL) from 2009 through 2011. Then, machine-learning methods for predicting (1) situations in which MIT restrictions for ATL arrivals are implemented under low demand scenarios, and (2) days in which a large number of MIT restrictions are required to properly manage and control ATL arrivals are presented. More specifically, these predictions were accomplished by using an ensemble of decision trees with Bootstrap aggregation (BDT) and supervised machine learning was used to train the BDT binary classification models. The models were subsequently validated using data cross validation methods. When predicting the occurrence of arrival MIT restrictions under low demand situations, the model was able to achieve over all accuracy rates ranging from 84% to 90%, with false alarm ratios ranging from 10% to 15%. In the second set of studies designed to predict days on which a high number of MIT restrictions were required, overall accuracy rates of 80% were achieved with false alarm ratios of 20%. Overall, the predictions proposed by the model give better MIT usage information than what has been currently provided under current day operations. Traffic flow managers can use these predictions to identify potential MIT restrictions to eliminate (e.g., those occurring during low arrival demand periods), and to determine the days in which a significant number of restrictions may be required.

CITATION: Wang, Y. and Grabbe, S., "Modeling Weather Impact on Airport Arrival Miles-in-Trail Restrictions," *SAE Int. J. Aerosp.* 6(1):2013, doi:10.4271/2013-01-2301.

1. INTRODUCTION

When the demand in a region of airspace or at an airport is expected to exceed the available capacity for an extended period of time, traffic flow management specialists within the Federal Aviation Administration (FAA) implement TMIs in order to mitigate these demand-and-capacity imbalances. Initiatives available to these specialists include Airspace Flow Programs, Ground Delay Programs, Ground Stops, rerouting and MIT restrictions. By far the most frequently used initiative is the MIT restriction, largely due to its

simplicity and ease of implementation [1]. Past surveys of Traffic Management Coordinators within the FAA indicate that MITs are most frequently fulfilled in response to volume and weather constraints [1]. Briefly, MIT restrictions are used to ensure a minimum spacing between aircraft in a stream, and they are typically implemented in increments of 5 nautical miles (nmi) in a range between 10 and 30 nmi. The number of aircraft entering a region of airspace controlled by a MIT restriction is inversely proportional to the MIT restriction.

In recent years, a number of MIT-related studies have appeared in the literature. For example, descriptions of currently practiced MIT operations and their impact on the National Airspace System (NAS) are presented in Refs. [1-2]. Since MIT restrictions are largely developed and carried out manually in current operations, some researchers proposed techniques for modeling the impact of MIT restrictions prior to operational implementation (see for example, [3-4]). In Ref. [3], a decision support capability, called the MIT Impact Assessment capability, which allows a user to model the impact of proposed MIT restrictions, is described. An analytic formulation and a linear programming approach for modeling the effects of MIT restrictions are introduced in Ref. [4]. Despite the past work in this area, there have been few published studies that seek to predict whether or not a MIT restriction will be required and if so what the value of the restriction should be. To the best of the authors' knowledge, the only work attempting to explore this issue is the one reported in Ref. [5]. In this study, the authors prototyped and evaluated four machine learning algorithms for providing probabilities that a particular MIT restriction may be applied to manage arrivals destined for Newark Liberty International Airport (EWR).

This work seeks to build on the work in Refs. [2, 3, 4, 5] by first offering an updated analysis of the cause, frequency and duration of historical MIT restrictions, and subsequently using machine learning techniques to predict the occurrence of MIT restrictions to manage arrivals into ATL airport. The historical analysis was accomplished using data from the FAA's National Traffic Management Log (NTML) database for 2009 through 2011 to obtain MIT start times, end times, locations, causes, and values. Since MIT restrictions often operate in conjunction with additional TMIs, records for Ground Delay Programs (GDP) and Ground Stops (GS) impacting ATL were also obtained. In addition to the NTML data, schedule arrival rates, actual arrival rates and arrival delays were obtained for 2009 through 2011 from the FAA's Aviation System Performance Metric (ASPM) system. Following the historical analysis, BDT was used to classify the weather-related MIT restrictions for ATL arrivals. Supervised machine learning was applied to train the BDT binary classification models, and the models were subsequently validated using data cross validation methods. The model was able to achieve overall accuracy rates ranging from 80% to 90% with false alarm ratios ranging from 10% to 20%.

The data mining and cross validation approach is described in Section 2. The NTML and ASPM data input sources are outlined in Section 3. The historical analysis of MIT restrictions is presented in Section 4, while the data mining predictions are presented in Section 5. Finally a summary of the results is presented in Section 6.

2. APPROACH AND MODELING METHODOLOGY

Data mining techniques were used to classify MIT restrictions in terms of variations in the arrival demand (i.e., demand less or greater than the airport arrival capacity) and the number of restrictions issued on a particular day (i.e., either a low number of restrictions or a high number of restrictions). In this case, an Ensemble Bagging Decision Tree model, described below, was used to classify the results, and supervised machine learning was applied to train the BDT binary classification models. Validation of the models was accomplished using data cross validation methods, as stated below.

Ensemble bagging decision tree

Ensemble methods use multiple machine learning decision tree models to obtain better predictive performance than what any of its individual constituent members can produce. Bagging stands for bootstrap aggregation. Bootstrap aggregation is a machine learning ensemble meta-algorithm designed to improve the stability and accuracy of machine learning algorithms used in statistical classification and regression [6]. In classification scenarios, the random resampling procedure in bagging induces some classification margin over the dataset. Additionally, when bagging is performed in different feature subspaces, resulting classification margins are likely to be diverse, which is essential for an ensemble to be accurate. This method takes into account the diversity of classification margins in feature subspaces to improve the performance of bagging. First, it studies the average error rate of bagging, converts the task into an optimization problem for determining some weights for feature subspaces. Then, it assigns the weights to the subspaces via a randomized technique in classifier construction. Experimental results demonstrate that the ensemble method is robust to classification noise and often generates improved predictions than any single classifier [7, 8, 9]. In this study, the BDT classification model is implemented using the MATLAB TreeBagger function.

Model validation methods

Machine learning 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 training data sets and have to be tested or validated using a different data set. Several validation approaches are available. Among them, the very popular one, which has been used frequently by researchers, is cross-validation.

In cross-validation, a series of BDT models are constructed, each time by dropping a different part of the data from the training set and applying the resulting model to the dropped data to predict the target. The merged series of predictions for dropped or test data are checked for accuracy against the observations. In one version of the cross-

Table 1. Confusion matrix for dichotomous (“Yes”/”No”) events.

		Actual Observation		Marginal Sum
		Yes	No	
Classifier Prediction	Yes	YY	YN	YY+YN
	No	NY	NN	NY+NN
Marginal Sum		YY+NY	YN+NN	YY+YN+NY+NN

validation approach, called group cross-validation approach, data are divided into N groups. A total of N models are then constructed one by one using N-1 data groups for model training, and the remaining group is used for testing. At the end of this procedure, all predictions assembled from the dropped cases are compared with the observed targets to compute validation of model error for the cross-validation result. The ten-fold cross-validation is used in this study.

Performance measures

A number of methods are available to evaluate the performance of binary classifiers. For a classifier with any given discrimination threshold, the number of cases correctly and incorrectly classified can be computed. This gives a confusion matrix with four numbers as shown in Table 1. YY is the number of true positives, i.e., how many cases are estimated by classifier as “Yes” events, which actually are “Yes” events. Similarly we can define NN as the number of true negatives, NY as the number of false positives and YN as the number of false negatives. Using the statistics generated in Table 1, some frequently used classifier performance evaluation methods are described briefly below. More information about these methods can be found in Refs. [9, 10, 11, 12].

The Overall Accuracy Rate (*OAR*) is defined as $OAR = (YY + NN) / (YY + YN + NY + NN)$. It has a range of 0 to 1. “1” is the best classification performance score. The probability of detection (*POD*), also called as precision, is the proportion of “Yes” observed events that were correctly predicted, $POD = YY / (YY + NY)$. The probability of false alarm (*PFA*), also called as false alarm ratio, is the proportion of “No” observed events that were not correctly estimated as “Yes” predicted events, $PFA = YN / (YY + YN)$. Its values also range from 0 to 1. If $YN = 0$, then the score goes to 0, the best one can expect. The Critical Success Index (*CSI*) is the proportion of true positives that were either estimated or observed. $CSI = YY / (YY + YN + NY)$. Its values range from 0 to 1 with a value of 1 indicating a perfect classification performance score. The *PFA* can be controlled by deliberately under-predicting the event; such a strategy risks increasing the number of missed events, which is not considered in the *PFA*. For this reason, the *POD* and the *PFA* should both be considered for a better understanding of the performance of the forecast.

The *OAR*, *POD*, *PFA*, and *CSI* classifier performance measures are used in this research.

3. EXPERIMENTAL SETUP

This section describes the TMI data, weather data, and air traffic data that were used in this study. The data sources are NTML [13], Center-level Weather Impacted Traffic Index (WITI) data [14], and ASPM database [15]. All data at MIT issue time for 2009 through 2011 were derived from these data sources.

MIT data

MIT start hours, actual durations and values (e.g., 10 MIT) were obtained from the NTML database and were used for historical analyses (see Section 4) and as the targets to the machine learning models (see Section 5).

GDP and GS data

GDP and GS start hours and actual durations were obtained from the NTML database. These records were primarily used for preconditioning the inputs to the machine learning models to identify time periods in which only MIT restrictions were impacting the ATL arrivals, as opposed to situations when both an MIT and a GDP or GS restriction were impacting the arrivals.

Airport terminal weather data

Actual hourly airport surface weather observations (METAR), such as wind, ceiling, visibility, and meteorological condition flags for the ATL airport were obtained from the ASPM database. These data were preprocessed to convert character records to numerical values and to filter out the missing ones. The processed METAR data were used as the inputs to the machine learning methods.

Center-level WITI data

The Center-level WITI data was calculated for the Atlanta Air Route Traffic Control Center (ARTCC) and used as a proxy to estimate the impact of en route convective weather on the ATL arrivals. Briefly WITI is an indicator of the number of aircraft affected by weather. The computation of the WITI values used in this study was accomplished by overlaying a one-degree latitude by one-degree longitude grid over the Atlanta ARTCC. The computation of WITI consists of (1) assigning a value of one to every grid cell W_{ij} of the weather grid W where severe weather is present and zero elsewhere, (2) counting the number of aircraft in every grid cell A_{ij} , and (3) computing the WITI as a function of time t using Eq. (1).

$$WITI(t) = \sum_{j=1}^m \sum_{i=1}^n A_{i,j}(t) W_{i,j}(t) \quad (1)$$

Air traffic data from the FAA's Aircraft Situation Display to Industry system was used to calculate $A_{i,j}$ while $W_{i,j}$ was calculated using current observations from the Convective Weather Avoidance Model [16].

Airport air traffic data

ATL arrival air traffic data during the MIT issuance times, such as scheduled arrival, the Enhanced Traffic Management System (ETMS) aircraft arrival counts, and arrival demand were collected from the ASPM database. The dynamic airport capacity, Airport Arrival Rate (AAR), was also gathered from the ASPM database for model development.

4. HISTORICAL ARRIVAL MILES-IN-TRAIL ANALYSIS

MIT restrictions can be used to control en-route (i.e., over flight) traffic, arrivals and departures. A preliminary analysis of the distribution of the historical MITs in Atlanta Center indicated that 23% of MITs were issued to control en-route traffic, 19% were issued for departures and 58% were issued to control arrivals in 2009. Because of the predominance of MITs issued for arrivals in this Center, they were the predominant restrictions considered in this study. In the remainder of this section, analyses of historical arrival MITs at ATL will be presented in terms of the cause, temporal usage (e.g., daily, weekly, monthly), duration and value, and usage in conjunction with other TMIs, such as GDPs and GSs.

4.1. Analysis of the Top-10 Airports

A distribution of the airport arrival MITs for the top 10 U.S. airports from 2009-2011 is given in Fig. 1. Chicago O'Hare International Airport (ORD) accounted for 26% of all arrival MIT restrictions, ATL accounted for 15% of the restrictions and Chicago Midway International Airport (MDW) accounted for 8% of the arrival MIT restrictions.

The major causal factors, as recorded in the NTML database, of the arrival MITs for the top-10 airports in the U.S. from 2009-2011 are shown in Fig. 2. As can be seen from this plot, "Volume" is the predominant stated cause for arrival MITs at most of the analyzed airports. The notable exception is for ORD where the cause for a large number of MITs was not specified. The presence of "Weather" related MITs at ATL, Los Angeles International Airport (LAX) and Denver International Airport (DEN) is also noteworthy, since they account for over 40% of all arrival MITs. In this figure, "Volume" is used to indicate air traffic congestion at arrival fixes or airports. Because of the prevalence of MITs impacting arrivals into ATL, the remainder of this section and

the following section will focus on the MIT restrictions impacting the ATL arrival flights only.

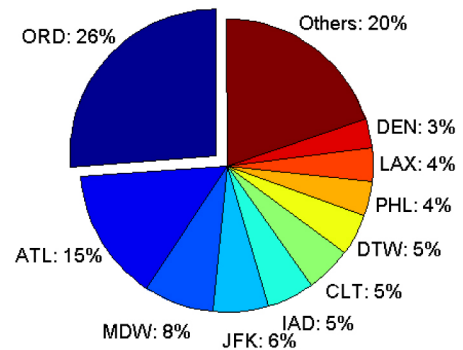


Figure 1. Distribution of Airport Arrival MITs from 2009-2011.

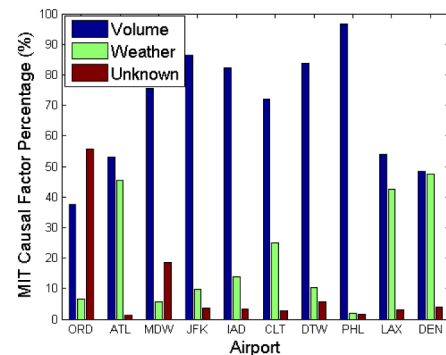


Figure 2. Causal Factors for arrival MITs at the top-10 U.S. airports from 2009-2011.

4.2. ATL Arrival MIT Statistics

The major weather causal factors associated with the arrival MIT restrictions at ATL from 2009-2011 are shown in Fig. 3. As indicated by the figure, low ceilings account for nearly 75% of all the restrictions, followed by thunderstorms, low visibility and winds. Weather phenomena, such as snow, rain and ice, accounted for less than 1% of causal factors, which is not surprising given the geographical location of ATL.

Temporal usage statistics (e.g., monthly, daily and hourly) for arrival MIT restrictions at ATL from 2009-2011 are exhibited in Fig. 4. The data is divided in terms of weather (blue bars) and non-weather (red bars) (labeled as "Volume & Unknown") events. Starting with the monthly usage statistics, which appear in the upper-most image in Fig. 4, it is noted that there is a slight decrease in arrival MIT restrictions in November and December timeframe. Additionally, there tends to be more weather-related MIT restrictions in the summer months (May through September), while "Volume & Unknown" related MIT restrictions are more prevalent in the winter and early-spring months. In terms of the weekly usage of arrival MIT restrictions at ATL (see the middle image in Fig. 4), the number of restrictions is