2013-01-2301
Published 09/17/2013
doi:10.4271/2013-01-2301
saeaero.saejournals.org

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 $W_{i,j}$ of the weather grid W where severe weather is present and zero elsewhere, (2) counting the number of aircraft in every grid cell $A_{i,j}$, and (3) computing the WITI as a function of time t using Eq. (1).

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

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.

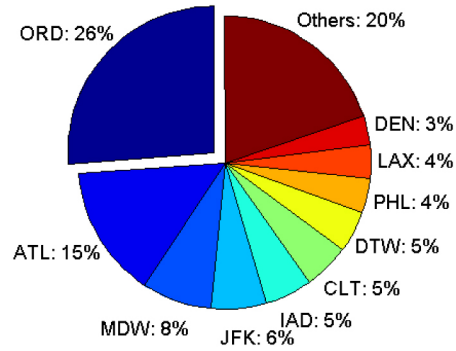

Figure 1. Distribution of Airport Arrival MITs from 2009-2011.

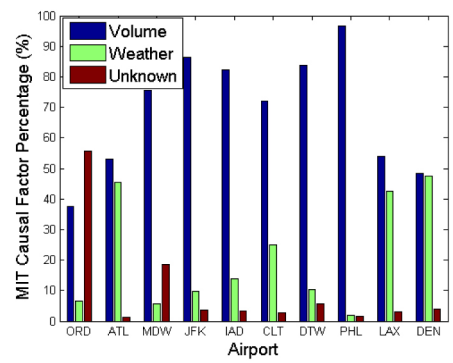

Figure 2. Causal Factors for arrival MITs at the top-10 U.S. airports from 2009-2011.

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

fairly constants but there is a noticeable decrease in the usage of these restrictions on Saturdays, which is to be expected since the arrival demand also tends to be lower on Saturdays. Finally, arrival MIT restrictions tend to be implemented early in the morning between 6:00am and 7:00 am local time (Easter Standard/Daylight Time), which coincides with the first bank of arrivals destined for the airport (see the bottom image in Fig. 4). There is also a second noticeable increase in arrival MIT restrictions in the late afternoon (16:00-18:00), which coincides with the last major bank of arrivals landing at the airport for the day.

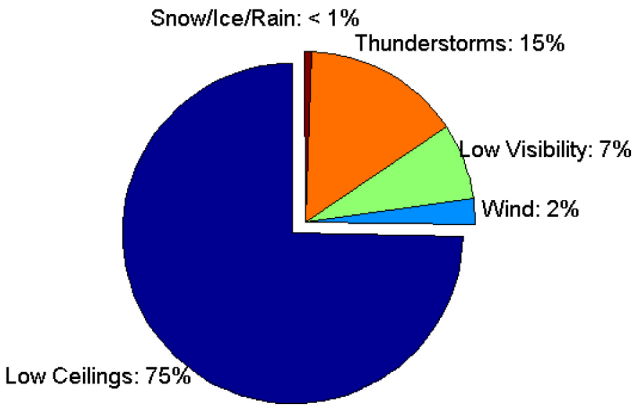

Figure 3. Weather Causal Factors for ATL arrival MITs from 2009-2011.

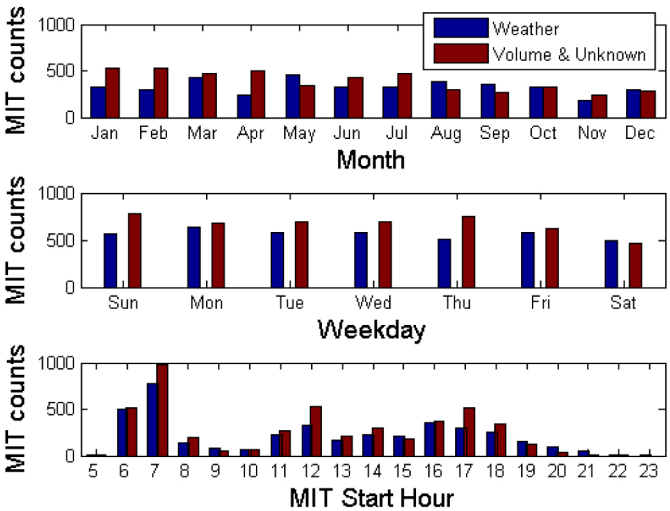

Figure 4. Temporal usage statistics for ATL arrival MIT restrictions from 2009-2011.

The duration and value of the arrival MIT restrictions for ATL from 2009-2011 are shown in Fig. 5. As stated in the previous analysis, the values are decomposed into weather and non-weather related events in this figure. Non-weather related restrictions tend to peak around 60 minutes, whereas a much larger spread in the duration is observed for weather related MITs. This seems logical since the "Volume & Unknown" restrictions are typically used to "smooth out" arrival pushes into the airport, which typically last on the order of an hour, while on the other hand, weather impacts at the airport are much more irregular, and can routinely last several hours. As can be seen from the lower part of Fig. 5, 20 MIT restrictions are most commonly issued for non-weather related events, whereas slightly higher restrictions (e.g., 25 MIT) are needed for weather related events.

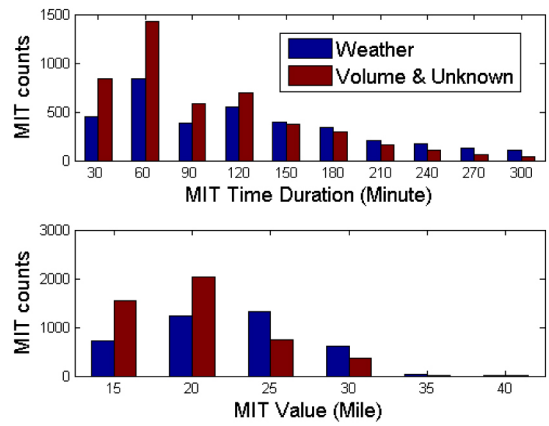

Figure 5. Histograms of Duration and Value of ATL arrival MIT restrictions from 2009-2011.

The arrival MIT restrictions are used to organize the air traffic flows at airport arrival fixes and protect the airport. The top five frequently used ATL arrival MIT restrictions are applied at the arrival fixes closest to ATL. The durations and the values of the arrival MIT restrictions at the closest arrival fixes to ATL from four directions are shown in Fig. 6. These closest arrival fixes to ATL are FLCON at the northeast (blue bars), CANUK at the southeast (cyan bars), LGC and HONIE at the southwest (yellow bars), and ERLIN at the northwest (magenta bars). The most MIT restrictions are implemented at fixes of LGC and HONIE which coincide with most arrival flights coming from the southwest (yellow bars in Fig. 6). Most restrictions tend to last around 60 minutes (see Fig. 6 a, 6 c); whereas a larger spread in the duration is observed for weather related MITs (Fig. 6 a). Most non-weather related events need 20 MITs(Fig. 6 d), whereas 25 MIT restrictions are frequently used for weather related events (Fig. 6 b) with an exception at ERLIN in which the most frequently used MIT restriction value from the northwest is 15 MITs (Fig. 6 b, 6 d).

The arrival MIT restrictions can be applied at several fixes at the same time. The ATL hourly arrival MIT restriction count percentages from local time 5 am to midnight for 2009 are listed in table 2. During these time period, 46% of the time is without using any ATL arrival MIT restrictions. The average of MIT hours (for hourly MIT counts>0) is about three restrictions per hour. Conceptually, the lower hourly MIT restriction counts, say one MIT per hour, could be used to intentionally smooth out air traffic congestion at arrival fixes, while the higher hourly MIT counts would be used to constrain air traffic from more

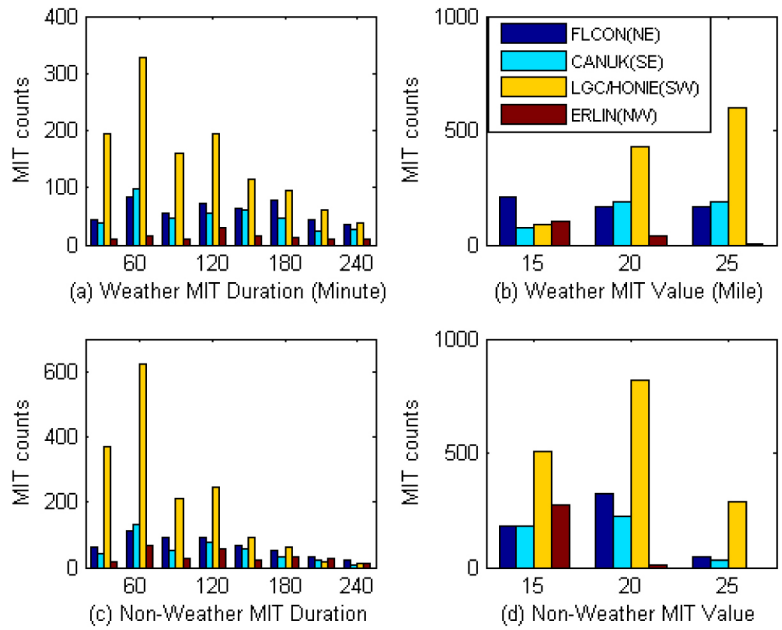

Figure 6. MIT Durations and Values for ATL MIT restrictions at arrival fixes from 2009-2011.

Table2. ATL hourly arrival MIT restriction counts

Hourly MIT Counts	0	1	2	3	4	5	6	>6
Percentage	46%	13%	10%	9.6%	9.0%	6.2%	3.5%	2.9%

directions destined towards the ATL airport to mitigate the imbalances of ATL arrival demand and ATL capacity.

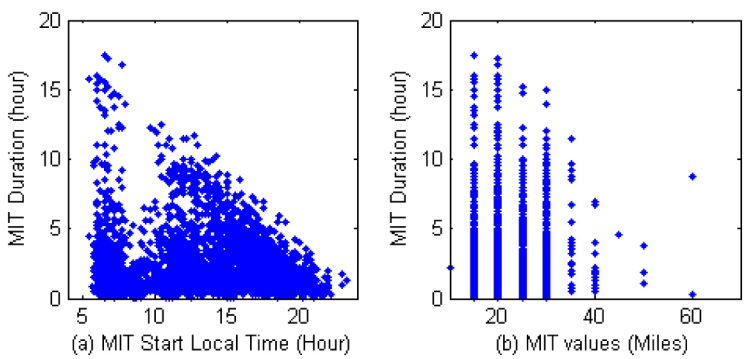

Figure 7. MIT Duration versus MIT Start Time (a) and MIT Value (b) for ATL arrival MIT restrictions from 2009-2011.

The MIT duration versus the MIT start time and the MIT duration versus the MIT value for all ATL arrival MIT restrictions from 2009 through 2011 are shown in Fig. 7. As would be expected, the MIT duration is strongly influenced by the MIT start time, and MIT restrictions implemented earlier in the day tend to be longer than those happened later in the day (see in Fig. 7 a). This trend strongly resembles the ATL arrival demand profile, which drops rapidly after 22:00 and remains low until the first "arrival push" that reaches ATL around 6:00am. The MIT duration results also exhibit a sharp decrease from roughly 8:00 am to 10:00 am, which coincides to the decrease in arrival demand between the major morning arrival rushes. On average, the longer duration MIT restrictions tend to have smaller values (e.g., less than 30 MIT) than the shorter duration MIT restrictions, as indicated by (b) in Fig. 7.

4.3. MIT Interactions with Ground Delay Programs and Ground Stops

Arrival MIT restrictions, which are local restrictions that typically control flights within Atlanta Center, are often operated in conjunction with national-level restrictions, such as Ground Delay Programs and Ground Stops, that are used to control all arrivals destined towards a capacity constrained airport. The rational being that the GDP/GS applies a course level of control, so that the appropriate number of arrivals reach an airport within a given hour, while the MIT restriction is used to "fine-tune" the spacing between individual flights in the arrival stream.

Figure 8 contains histograms of the hourly arrival MIT restriction counts over which ATL arrival MIT restrictions operate in the presence of both ATL GDP and GS (see in Fig. 8 a), absence of ATL GDPs (Fig. 8 b), absence of GS (Fig. 8 c), and absence of both GDP and GS (Fig. 8 d) respectively. When an arrival MIT restriction operated in conjunction with a GDP or a GS, it was most common with around 3 MIT restriction counts per hour (Fig. a, b, c). In contrast when

arrival MIT restrictions operate independently of GDPs and GSs, typically only one restriction was applied, as indicated by the (d) in Fig. 8.

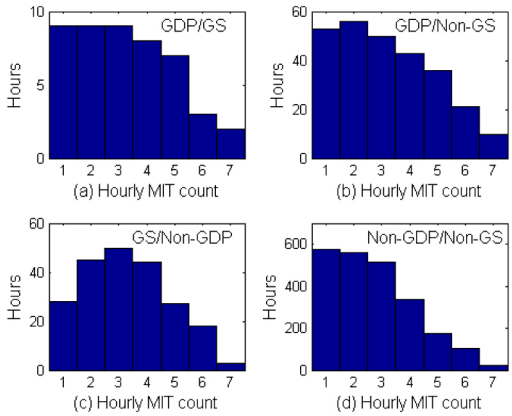

Figure 8. Hourly ATL arrival MIT restriction counts implemented with both GDP and GS (a), with GDP without GS (b), with GS without GDP (c) and without both GDP and GS (d) for the year 2009.

4.4. MIT Influence on Arrival Delays

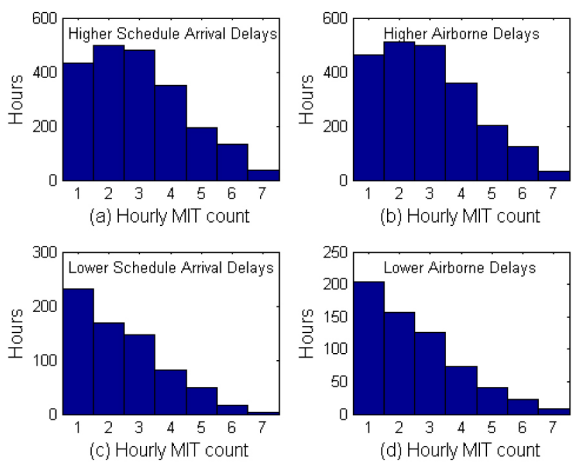

Figure 9. Influence of hourly ATL arrival MIT counts on schedule arrival and airborne delays for the year 2009.

Conceptually, higher hourly arrival MIT restriction counts should lead to higher delays for the ATL arrivals. To test this, the 2009 through 2011 ATL arrival MIT data was partitioned into two sets based on whether the arrival delays during a particular hour were either greater or less than the median delays for ATL arrivals for the year 2009. The results are presented in Figure 9 where the histograms (a) and (c) present the hourly arrival MIT restriction counts for schedule arrival delays and (b) and (d) for airborne delays. Fig. 9 (a) and 9 (b) reveal the hourly ATL arrival MIT restriction counts when the delays are greater than the median of the delays, and the lower parts (Fig. 9 c, d) show the corresponding results when the delays are less than the median delays. When the arrival delays are less than the median arrival delays, the most frequently used hourly arrival MIT restriction counts is one (Fig. 9 c, d). However, when the arrival delays are greater than the median arrival delays, there is tendency for hourly arrival MIT restriction counts over two (Fig. 9 a, b).

4.5. MIT Variations with Arrival Demand and Arrival Counts

Since arrival MIT restrictions are used to enforce a minimum spacing between subsequent flights in an arrival stream, it is expected that the usage of MIT restrictions should increase as the arrival demand approaches the airport capacity and decrease as the number of flights in the arrival stream decreases. To test this hypothesis, two sets of analyses were undertaken. In the first analysis, the hourly ATL arrival demand statistics from the year 2009 were partitioned into two sets, and the hourly MIT restriction counts in each set were investigated (see Figure 10 a, c). The first set (Fig. 10 a) of data held all hourly records where the arrival demand was greater than 80% of the AAR, whereas in the second set (Fig. 10 c), the demand was less than or equal to 80% of the AAR. As would be expected, typically hourly arrival MIT restriction counts from the first set were higher than the counts from the second set. For the hours during which the demand was greater than 80% of the AAR, the hourly MIT restriction counts were typically over two restriction counts (Fig. 10 a), while the hourly MIT counts were typically in place of one MIT per hour as displayed in Figure 10 (c).

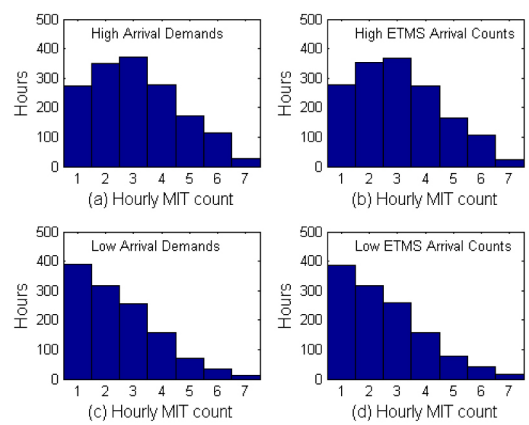

Figure 10. The relations between and ATL hourly Arrival MIT Restriction counts and ATL hourly arrival demand and actual arrival counts for the year 2009.

To determine if the MIT usage trends vary with the actual arrival counts, the same trial was repeated by partitioning the hourly arrival records into the following two sets: one with ETMS arrival counts greater than the median of ATL ETMS arrival counts (see in Fig. 10 b) and the other one with ATL ETMS arrival counts less than the median ETMS arrival count (Fig. 10 d). The comparison between (b) and (d) in Fig. 10 clearly reveals that the hourly applied MIT counts were

strongly influenced by the actual arrival counts. In both analyses, higher hourly MIT restriction counts were associated with either the higher demand or higher actual arrival counts at ATL as shown in Fig. 10.

In the results presented in Fig. 10, no effort was made in the hourly records to distinguish if a GDP/GS was being operated in conjunction with a MIT restriction or not, but as was discussed in Section 4.3 it is fairly common for an arrival MIT restriction to be implemented in conjunction with an ATL GDP or GS. To extend the work presented in Figs. 8 and 10, the hourly normalized arrival demand and the hourly normalized actual available arrival airport capacity as a function of the applied restrictions (e.g., GDP/GS, MITs, or no TMI restrictions) were displayed in Fig.11. The normalized arrival demand is defined as hourly demand/ AAR-1, and the normalized actual available arrival airport capacity is defined as 1-ETMS Arrival Counts/AAR.

Beginning with the results presented in Fig. 11 (a), (c) and (e), the normalized hourly arrival demand is seen to vary between −1 and two. A value of zero indicates situations in which the arrival demand is equal to the AAR, negative values signal those situations where the demand is less than the AAR, and positive values occur when the demand exceed the AAR. It should be noted that the vertical axis in Fig. 11 is the event hours observed over the year 2009. Observations associated with these three cases are stated as below:

• It is apparent that a GDP or GS is implemented when the arrival demand exceeds the AAR (e.g., demand/AAR −1 > 0) (Fig.11 a);

• When the demand is near or slightly below the AAR, there is a tendency to use arrival MIT restrictions alone without GDPs or GSs (Fig. 11 c). This shows a potential area for improvement in establishing effective airport arrival MIT controls because, from the perspective of an airport traffic operation, using arrival MIT restrictions to slow down arrival traffic flow is not needed if the airport arrival demand is less than airport arrival capacity;

• Lastly, when the demand is significantly below the AAR (Fig. 11 e), neither MIT nor GDP/GS is typically required to control the arrival traffic flows.

Figure 11 (b), (d) and (f) present a complimentary set of images in terms of the normalized hourly actual available arrival airport capacity (1 - ETMS Arrival Counts/AAR). In contrast to the results presented in Fig. 11 (a), (c) and (e), a value of zero in this case indicates that the hourly actual arrival counts were equal to the AAR, namely there is no space for additional arrival aircraft to land, and positive values indicate that there is a room for landing. A pattern can be summarized from Fig. 11 as follows:

• A tendency to control ATL arrivals with a GDP or GS when the actual arrival counts are slightly less than the AAR, i.e. the normalized available capacity is less than 20% of AAR (Fig. 11 b);

• A trend to control ATL arrivals only with MIT restrictions when the normalized available capacity is less than 50% of AAR (Fig. 11 d);

• When the hourly arrival counts are significantly less than the AAR, no MIT or GDP/GS is required to control the arrival traffic flows (Fig. 11 f).

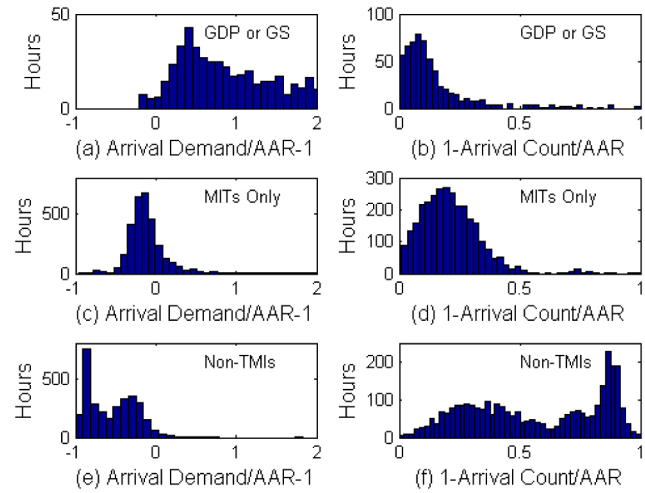

Figure 11. ATL hourly Demand/AAR-1 (a, c, e) and 1-ETMS Arrival Count/AAR (b, d, f) distributions in the Presence of GDPs/GSs, MITs, and Absence of TMIs for the year 2009.

As mentioned before, there is possibility that the usage of arrival MIT restrictions is not necessary when the airport arrival demand is less than airport AAR. Reducing the number, duration and value of ATL MIT restrictions under these conditions may allow better utilization of the airport and avoid unnecessary delays. In the remainder of this study, machine learning methods for optimizing the usage of MIT restrictions were presented and validated. The methods may have potential to be used by the traffic flow specialists to aid in determining whether or not an MIT restriction is necessary to control arrivals destined for ATL.

5. CLASSIFICATION RESULTS

This section contains the classification results generated using an Ensemble Bagging Decision Tree model to (1) examine the usage of arrival MIT restrictions under low arrival demand scenarios and to (2) estimate MIT usage on days that historically required a significant number of MIT restrictions. The former of these two studies was designed to identify situations in which arrival MIT restrictions may have been unnecessary, last too long or have a value that was too high, while the later was to aid the specialists in their daily post-operation review to evaluate if the restrictions implemented in that day are effective or not based on the model. In both cases, supervised machine learning was used to train the BDT binary classification models, and model validation was accomplished with ten-fold cross validation.

5.1. MIT Classification for Low Arrival Demand Scenarios

For predicting if MITs were used to control ATL arrivals under low demand situations, the arrival MIT training data was grouped into two classes. The first class was labeled "Yes", and was used to indicate that an arrival MIT was used under a low demand scenario, while the second was labeled "No" to indicate arrival MIT usage under high demand scenarios. Using the binary indicator responses of arrival MIT usage as targets, the BDT classification models were first trained, and subsequently applied to the test data for prediction purposes. Two alternative approaches were examined to determine when a low demand scenario occurred - the first involved analyzing the traffic demand, actual traffic counts and airport arrival rate for the duration of the MIT restriction, while the second involved examining the traffic demand and arrival rate on an hourly basis. The former of these approaches will be explored in more detail in Section 5.1.1, and the later will be examined in Section 5.1.2.

5.1.1. Restriction-based classification

In the restriction-based approach, all of the hours associated with an arrival MIT restriction are examined to identify situations in which an arrival MIT restriction was implemented during a low demand and low arrival count situation. Identifying these situations is accomplished using the following technique.

Given a set of arrival MIT restrictions M, intermediate variables α_m ($\alpha_m \in \alpha$) and β_m ($\beta_m \in \beta$) of a MIT restriction m ($m \in M$) can be defined as

$$\alpha_m = \min\left(\{f(t_m) : t_m \in T_m\}\right) \tag{2}$$

and

$$\beta_m = \max(\{g(t_m) : t_m \in T_m\}) \tag{3}$$

Where the normalized actual available arrival airport capacity $f(t_m) = 1 - C(t_m)/AAR(t_m)$, the normalized arrival demand $g(t_m) = D(t_m)/AAR(t_m) - 1$ and the MIT time set of m $T_m = \{T_{1,m},\ldots,T_{n,m}\}$. $C(t_m)$, $D(t_m)$ and $AAR(t_m)$ are the corresponding actual airport arrival count, airport arrival demand, and airport arrival rate at time t_m, respectively. The t_m is the time (date and hour) between the start ($T_{1,m}$) and the end ($T_{n,m}$) time that the MIT restriction m was implemented. The set of MIT restrictions with low demand and high available capacity, called as "low demand" in short, is then specified as

$$S = \{m : \alpha_m > 0.1 \cup \beta_m < 0\} \tag{4}$$

By this definition, 1,345 out of the 3,079 total MIT arrival records for the year 2009 (where there were no GDPs or GSs) were detected as low demand MIT restrictions. These arrival MIT restrictions were labeled as "Yes" and the remaining MITs were labeled as "No". These binary values were treated as the target variables in training and validating the model. The terminal weather and airport conditions discussed in Section 3 were applied as the inputs to the classification. These inputs were selected at one hour before the MIT start times, since the ATL arrival MIT restriction approval time is, on average, 55 minutes earlier then the MIT start time for the year 2009. The scatter plot of α versus β is shown in Fig. 12. The red lines in Fig. 12 separates the whole space into two sets with the upper-left representing the set for the MIT restrictions implemented during the low demand period, while the remaining space represents the high demand or low available capacity period. The set S consists of those points enclosed in the up-left portion of Fig. 12.

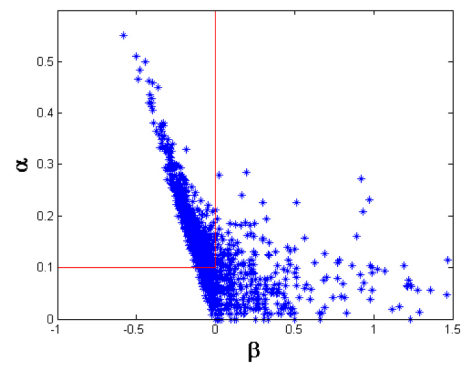

Figure 12. The α vs. β values for all ATL arrival MIT restrictions during 2009.

The 1 hr prediction of low demand MIT restrictions at ATL is shown in Table 3. Of the 3,079 total MIT events, 1,345 qualified as low demand events. The accuracy of the BDT binary classifier, which is given by OAR, is the proportion of correct results, (1161+1606)/(3079) = 0.90. Out of a total of 1,345 predicted low demand MIT events, the number of correctly predicted events was 1,161. The precision is then given by 1,161/1,345= 0.86 (see POD in Table 3).

The scatter plot of α versus β for all MIT restriction predictions is shown in Fig. 13. The red dots and blue dots in Fig. 13 are used to indicate the predictions for low demand and high demand MITs, respectively. It illustrates that most MIT restrictions are correctly predicted as the MIT implemented during the low and high demand periods. Overall, the BDT model was very good in identifying ATL arrival MIT restrictions that were being implemented during low demand periods. A review of the MIT restrictions implemented under these conditions may help to eliminate or reduce potentially unnecessary restrictions, MIT durations or values.

Table 3. 1-hr Forecast of Low Demand/Arrival ATL Arrival MIT Restrictions.

1 hr BDT Prediction		Actual Observation		
		Yes	**No**	**Marginal Sum**
Classifier Prediction	**Yes**	1161	128	1289
	No	184	1606	1790
	Marginal Sum	1345	1734	3079
OAR: 90%, POD: 86% PFA: 9.9%, CSI: 0.79				

Table 4. 1-hr Prediction of Low Demand ATL Arrival MIT Restriction hours.

2 hr BDT Prediction		Actual Observation		
		Yes	**No**	**Marginal Sum**
Classifier Prediction	**Yes**	1321	239	1560
	No	260	1313	1573
	Marginal Sum	1581	1552	3133
OAR: 84%, POD: 84% PFA: 15%, CSI: 0.73				

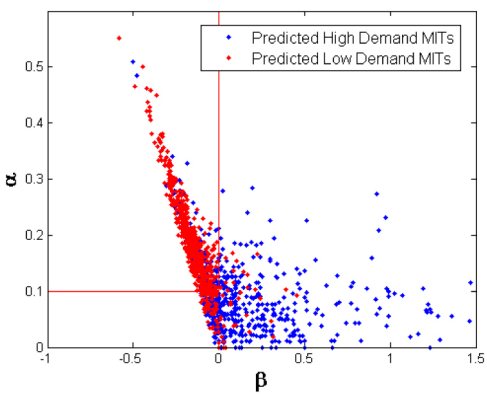

Figure 13. The α vs. β values for ATL arrival MIT restrictions predictions during 2009.

5.1.2. Hourly low demand classification

This is an extension of the results presented in Section 5.1.1. In this section, the hourly records associated with all arrival MIT restrictions are examined. As opposed to the results presented in Section 5.1.1, which was identifying potential opportunities for eliminating the usage of entire low demand MIT restrictions, this extension may be useful in determining if the time durations (hours), for example, of one or more restrictions that were implemented under a low demand scenario could be reduced. For this effort, the set of all low demand MIT hours is given as follow:

$$Z = \{g(t) < -0.15 : t \in T\} \tag{5}$$

Here $g(t)$ is defined as in Section 5.1.1 and the ATL MIT restriction hour set $T = T_1 \cup T_2 \cup \ldots T_m \cup \ldots$ includes all MIT restriction hours. Note that in contrast to the definition in Eq. (4), the definition in Eq. (5) does not depend on the hourly normalized actual available arrival airport capacity, $f(t)$. The scatter plot of $f(t)$ versus $g(t)$ is shown in Fig. 14. Using the 2009 hourly MIT arrival records for ATL (where there were no GDPs or GSs), a total of 1,581 of the hourly records out of 3,133 MIT records were determined to be low demand MIT hours. These events are graphically depicted as the set of all points laying at the left side of the red vertical line in Fig. 14.

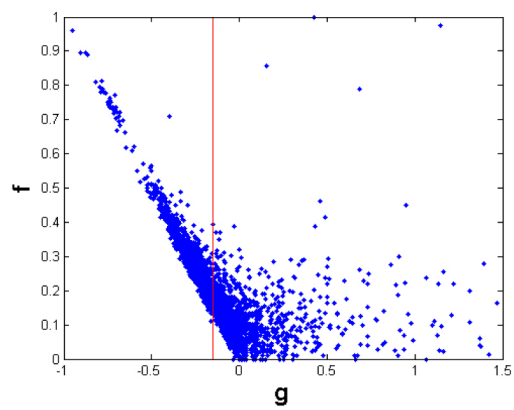

Figure 14. The hourly values f(t) versus g(t) of all ATL arrival MIT restriction hours for the year 2009.

The 1 hr predictions of low demand arrival MIT hours at ATL are listed in Table 4. As before a value of "Yes" is used to indicate an hourly low demand arrival MIT event, while a value of "No" indicates a high demand hourly event. The accuracy of the BDT binary classifier, which is given by OAR, was 84% for the 1-hr prediction (see Table 4). The precision for the 1-hr forecast was 84% (see POD).

The scatter plot of $f(t)$ versus $g(t)$ for all hourly MIT predictions is shown in Fig. 15. The red dots and blue dots in Fig. 15 are used to indicate the predictions for low demand and high demand MIT hours, respectively. Most MIT restriction hours are correctly predicted as those actually implemented during the low and high demand hours. As the results exhibited in Section 5.1.1, the BDT model was able to reliably recognize those arrival MIT restrictions being

Table 5. BDT Prediction of High MIT Usage Days for ATL using MIT Counts.

BDT Prediction		Actual Observation		
		Yes	**No**	**Marginal Sum**
Classifier Prediction	**Yes**	379	95	474
	No	129	492	621
	Marginal Sum	508	587	1095
OAR: 80%, POD: 75% PFA: 20%, CSI: 0.63				

implemented during low demand hours. These predictions may be useful to guide operation decisions to reduce potentially unnecessary hours or values for the MIT restrictions.

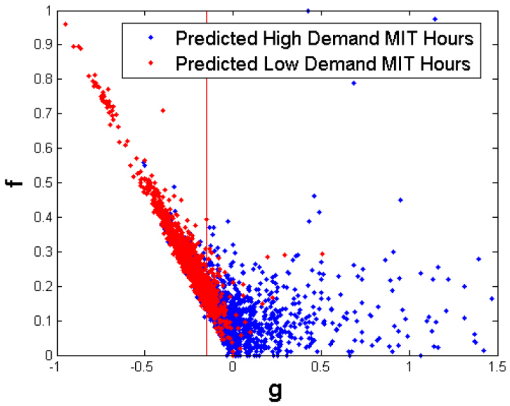

Figure 15. The hourly values f(t) vs. g(t) for ATL arrival MIT restrictions predictions.

5.2. MIT Classification for High Restriction Usage Scenarios

The ability to predict days requiring a significant number of MIT restrictions could aid traffic flow specialists in preparing for high workload situations. In order to predict if a large number of restrictions is required or not on a given day, the arrival MIT data was grouped into two classes. The first class labeled as "Yes" was used to indicate that a high number of arrival MIT restrictions was required on a particular day, while the second one labeled as "No" was to signify that a low number of MIT restrictions was needed on that day. Using the binary indicator responses of arrival MIT usage as targets, the BDT classification models were first trained, and then subsequently applied to the test data for prediction purposes. Two alternative approaches were examined in determining if a day required a high number of MIT restrictions or not. The first one relied on investigating the total number of arrival MIT restrictions implemented on a given day, while the second examined the product of the MIT durations and the values over that day. For both approaches, daily Atlanta Center WITI (see Section 3) values, daily average airport weather conditions, daily average scheduled arrival rates, actual arrival counts and airport arrival rates from 2009 through 2011 were used as inputs to the model. The first approach for identifying days with large number of restrictions will be described in more detail in Section 5.2.1, while the second one will be described in Section 5.2.2.

5.2.1. Daily MIT counts

To identify high arrival MIT usage days at ATL, the daily MIT counts from 2009 through 2011 were calculated from available NTML data. A histogram of the 1,095 daily records is shown in Fig. 16 where the red vertical line indicates the median value of five. For classification purposes, days with more than five arrival MIT restrictions were classified as high restriction days (labeled as "Yes"), and days with fewer were labeled as "No".

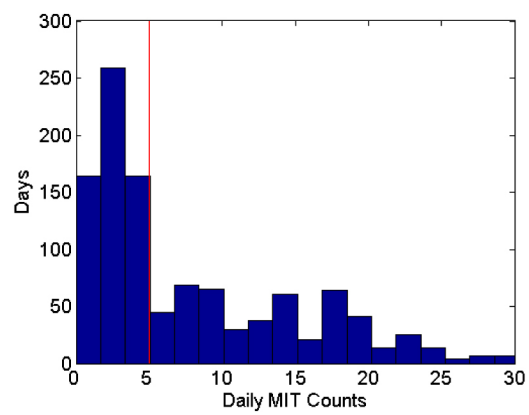

Figure 16. Daily MIT Restriction Counts for ATL Arrivals from 2009-2011.

The BDT prediction of high daily arrival MIT usage at ATL is shown in Table 5. Of the 1,095 total MIT hourly events, 508 qualified as high daily usage events. The accuracy of the BDT binary classifier, which is given by OAR, is the proportion of correct results, (379+492)/(1095) = 0.80. Out of a total of 474 predicted high MIT usage events, the number of correctly predicted events was 379. The precision is then given by 379/508= 0.75 (see POD in Table 5). Overall, the BDT model accurately identifying days in which a large number (e.g., more than five) of MIT restrictions were being used to control arrivals into ATL. As previously mentioned, an ability to correctly predict the occurrence of these types of days, may help traffic flow specialists by identifying these potentially high workload situations.

Table 6. BDT Prediction of High MIT Usage Days for ATL using MIT Durations and Values.

BDT Prediction		Actual Observation		
		Yes	**No**	**Marginal Sum**
Classifier Prediction	**Yes**	433	99	532
	No	114	449	563
	Marginal Sum	547	548	1095
OAR: 81%, POD: 79% PFA: 19%, CSI: 0.67				

5.2.2. Daily MIT durations and values counts

A second approach for identifying days with a high number of arrival MIT restrictions involves calculating the following expression for each day:

$$\delta_d = \sum_{i=1}^{N_d} D_{i,d} * V_{i,d} \tag{6}$$

Here $D_{i,d}$ is the duration of the ith arrival MIT restriction that was implemented on day d, $V_{i,d}$ is the value of the MIT restriction and N_d is the number of restrictions implemented on day d. The values of δ_d from 2009 through 2011 are shown in Fig. 17. The red vertical line in Fig. 17 is used to identify the median value of δ_d, which is equal to 189 hour*mile. For classification purposes, days with $\delta_d > 189$ were categorized as high restriction days (labeled as "Yes"), and days with fewer were labeled as "No".

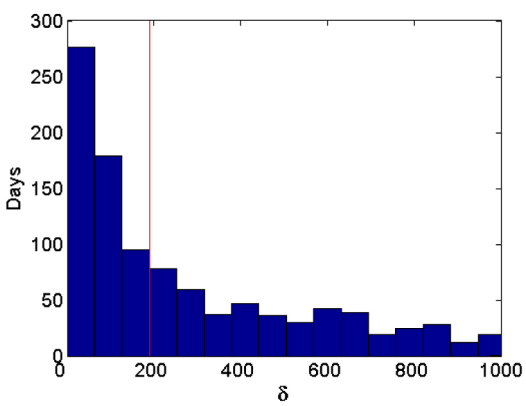

Figure 17. Sum of the product of the MIT values and durations for ATL Arrivals MIT Restrictions from 2009-2011.

The BDT prediction of high daily arrival MIT usage at ATL is shown in Table 6. Of the 1,095 total MIT hourly events, 547 qualified as high daily usage events. The accuracy of the BDT binary classifier, which is given by OAR, is the proportion of correct results, (433+449)/(1095) = 0.81. Out of a total of 532 predicted high MIT usage events, the number of correctly predicted events was 433. The precision is then given by 433/547= 0.79 (see POD in Table 6). Once again, the BDT model accurately identified the days in which high δ values of MIT restrictions were being used to control the arrivals into ATL. As stated before, this ability to correctly predict the occurrence of these types of days may help traffic flow specialists by identifying these potentially high workload situations.

6. CONCLUDING REMARKS

This paper begins by first providing an extensive analysis of the MIT restrictions impacting ATL arrivals from 2009 through 2011. The paper subsequently presents 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. These predictions were accomplished by using an Ensemble Bagging Decision Tree (BDT) and supervised machine learning was employed to train the BDT binary classification models. The models were 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 that were 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%.

In summary, the predictions proposed by the model give better MIT usage information than what is being done under current day operations. Traffic flow managers may use these predictions in their decision making to determine potential MIT restrictions to eliminate (e.g., those occurring during low arrival demand periods), and identify days in which a significant number of restrictions may be required.

There is a room for improving the models described in this paper. For example, the estimates for other factors used in MIT planning, such as the MIT value, duration and impacted arrival fix, would be useful in developing better models. Additionally, longer duration predictions (e.g., 3-4 hours) would also make these models more operationally relevant.

REFERENCES

1. Kopardekar, P., Green, S. and Roherty, T., "Miles-in-Trail Operations: a Perspective," 3rd AIAA Aviation, Technology, Integration, and Operations (ATIO) Conference, 2003.
2. Myers, T.A., Klopfenstein, M., Mintzer, J., Wilmouth, G., Sud, V., "A Preliminary Analysis of the Impact of Miles-in-Trail Restrictions on NAS Flight Operations," 6th USA/Europe Air Traffic Management Research and Development Seminar, 2005.

3. Ostwald, P., Topiwala, T. and DeArmon, J., "The Miles-in-Trail Impact Assessment Capability," 6th AIAA Aviation, Technology, Integration, and Operations (ATIO) Conference, 2006.
4. Grabbe, S. and Sridhar, B., "Modeling and Evaluation of Miles-in-Trail Restrictions in the National Airspace System," AIAA Guidance, Navigation and Control Conference, 2003.
5. Bloem, M., Hattaway, D. and Bambos, N., "Evaluation of Algorithms for a Miles-in-Trail Decision Support Tool," Proc. of the International Conference on Research in Air Transportation, Berkeley, CA, 2012.
6. Breiman L., "*Bagging Predictors*," *Machine Learning*, vol. 24, no. 2, pp. 123-140, 1996
7. Dietterich T.G., "Ensemble Methods in Machine Learning," *Proc. Conf. Multiple Classifier Systems*, pp. 1-15, 2000.
8. Melville P., Shah N., Mihalkova L., and Mooney R.J., "Experiments with Ensembles with Missing and Noisy Data," *Proc Fifth Int'l Workshop Multiple Classifier Systems*, pp. 293-302, 2004.
9. Wang Y., "*Prediction of weather impacted airport capacity using ensemble learning*", in Proceedings of the 30th AIAA/IEEE Digital Avionics Systems Conference (DASC), October 2011.
10. 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.
11. "*Accuracy and Precision*." n.p., n.d. Web. July 26, 2011. <http://en.wikipedia.org/wiki/Accuracy_and_precision>.
12. Fukunaga, K., "*Introduction to Statistical Pattern Recognition*", Academic Press, 1990.
13. "*Air Traffic Organization Policy*." FAA, Feb 11, 2010. Web. July 26, 2011. <http://www.faa.gov/air_traffic/publications/atpubs/FAC/1705.html>.
14. Sridhar, B. and Swei, S.S.M., "Relationship between Weather, Traffic and Delay Based on Empirical Methods," 6th AIAA Aviation, Technology, Integration, and Operations (ATIO) Conference, 2006.
15. "*FAA Operations and Performance Data*." FAA, n.d. Web. July 26, 2011 <http://www.apo.data.faa.gov/>.
16. Evans, J.E. and Ducot, E.R., "Corridor Integrated Weather System," Lincoln Laboratory Journal, Vol. 16, No. 1, pp. 59-80, 2006.

DEFINITIONS/ABBREVIATIONS

AAR - Airport Acceptance Rate or Airport Arrival Rate

ARTCC - Air Route Traffic Control Center

ASPM - Aviation System Performance Metric system

BDT - An ensemble Bagging Decision Trees. Bagging stands for bootstrap aggregation. A decision tree is a decision support tool that uses a tree-like model of decisions and their possible consequences

CSI - Critical Success Index

ETMS - Enhanced Traffic Management System

FAA - Federal Aviation Administration

GDP - Ground Delay Program

GS - Ground Stop

METAR - Meteorological Terminal Aviation Routine Weather Report

MIT - miles-in-trail restrictions

NAS - National Airspace System

NTML - National Traffic Management Log

OAR - Overall accuracy rate

PFA - Probability of false alarm

POD - Probability of detection

TMI - Traffic management initiatives

WITI - Weather Impacted Traffic Index