Clustering Days with Similar Airport Weather Conditions

Shon Grabbe[1] and Banavar Sridhar[2]
NASA Ames Research Center, Moffett Field, CA, 94035-1000

Avijit Mukherjee[3]
University of California Santa Cruz, Moffett Field, CA, 94035-1000

On any given day, traffic flow managers must often rely on past experience and intuition when developing traffic flow management initiatives that mitigate imbalances between aircraft demand and weather impacted airport capacity. The goal of this study was to build on recent efforts to apply data mining classification and clustering algorithms to vast archives of historical weather and air traffic data to identify patterns and past decisions that can ultimately inform day-of-operations decision-making. More specifically, this study identified clusters of hours for which the probability of imposing a Ground Delay Program was similar. The identification of these similar hours was a two-step process. In the first step, historical weather data, scheduled arrival rates and traffic flow management logs were used to estimate the hourly probability of a Ground Delay Program being implemented at both Chicago O'Hare International Airport and Newark Liberty International Airport. These hourly probabilities were subsequently clustered in the second-step to identify collections of hours in which the likelihood of a Ground Delay Program was similar. For Chicago, a total of five clusters were identified, while thirteen clusters were identified for Newark. This large difference reflects the increased diversity in air traffic and weather conditions that result in Ground Delay Programs being implemented at Newark, as compared to Chicago. For both airports, clusters associated with nighttime and early morning operations were clearly identifiable, and the probabilities associated with Ground Delay Programs occurring for these clusters was very low, as expected. Additional clusters representing typical "bad weather" days were also identified where the probabilities of a Ground Delay Program being implemented was high, but these were rare and accounted for less than 3% of all the hourly events. Somewhat surprising was the large number of clusters identified for Newark where the weather and traffic conditions were similar, but very inconsistent Ground Delay Program usage was observed. Finally, the frequency of occurrence of Ground Delay Programs, Ground Stops and Miles-in-Trail restrictions implemented for each of the clusters was analyzed. Based on the results, it does appear as if the usage of Miles-in-Trail, Ground Delay Program and Ground Stop restrictions correlates well with the probabilities of Ground Delay Program occurrence. Furthermore, the results demonstrate that it is feasible to use historical weather and air traffic archives to provide guidance on the types of traffic management initiatives to implement in response to the weather and traffic conditions impacting an airport.

I.   Introduction

WHEN severe weather impacts the airspace in and around an airport, traffic flow managers implement Traffic Management Initiatives (TMIs), such as Ground Delay Programs (GDPs), Ground Stops (GSs), Airspace Flow Programs (AFPs), Miles-in-Trail (MIT) restrictions and reroutes, in order to mitigate imbalances between the

---

[1] Research Scientist, Systems Modeling & Optimization Branch, Mail Stop 210-15, shon.grabbe@nasa.gov, AIAA Associate Fellow.
[2] Senior Scientist, Aviation Systems Program, Mail Stop 210-10, AIAA Fellow.
[3] Associate Project Scientist, University of California Santa Cruz, Mail Stop 210-8, AIAA Member.

aircraft demand and the airport capacity. Decision support tools, such as the Flight Schedule Monitor, are available to traffic management coordinators (TMCs) for implementing some of these restrictions, but information about how similar weather phenomena were handled in the past are often left to the decision makers' past experience and intuition, which are often collectively discussed during the strategic planning teleconferences that operationally occur roughly every two-hours. Fortunately, the controls implemented by the TMCs have been digitally archived in the National Traffic Management Log (NTML)[1] since the mid-2000s. This study expands on recent efforts to leverage these historical archives to identify patterns and trends at the airport-level with the aim of initially supporting post-operations analyses and ultimately day-of-operations decision-making.

Prior work that has focused on identifying similarities in the weather and traffic patterns at the airport-level includes the works of Klein[2] and Grabbe et al.[3] In Klein[2], a "day-at-a-glance weather impact matrix" for visualizing similarities of weather impacts both at the NAS-level, as well as the airport level, was proposed. Klein's methodology relied on the extension of the original en route Weather Impacted Traffic Index (WITI)[4] concept to the terminal area in order to account for airport-level weather impacts, such as reduced ceilings. Similar days were subsequently calculated by characterizing each day in terms of a vector of N-attributes (or features), and the standard cosine similarity measure from the machine learning domain[5] was used to identify similarities. Although Klein's approach would have been ideal for supporting the current study, the exact methodology for calculating the terminal WITI parameters was not well documented in the publically available literature, so an alternative approach was pursued in this study that built on the original work of Grabbe et al.[3] In Grabbe et al.[3], the weather conditions associated with historically implemented GDPs, as recorded in the NTML, were used to identify 18 similar weather days in the NAS from an airport perspective. Although this approach bears merit from a post-operations point of view, where the impacting weather conditions are known a priori, this technique is difficult to extend to day-of-operations planning since it does not explicitly take into account the actual weather and traffic conditions.

This work extends the work of Grabbe et al.[3] by directly utilizing weather observations and scheduled arrival traffic demand estimates to identify similar days at Newark Liberty International Airport (EWR) and Chicago O'Hare International Airport (ORD). These two airports were selected for testing based on the high frequency of GDP usage[6] and for the conditions for which the GDPs are typically implemented. For reference, EWR had the highest cumulative number of GDPs from 2007 through 2009[6], and these GDPs were generally associated with frontal passages that can lead to increased winds, precipitation, low ceilings and poor visibility. Chicago O'Hare International Airport on the other hand had the fifth highest number of GDPs from 2007 through 2009, but these GDPs were implemented for a wide range of conditions, such as snow/ice, winds, thunderstorms and low ceilings.[6] The first step in identifying similar days at EWR and ORD entailed using 2012 hourly Localized Aviation MOS (Model Output Statistics) Program (LAMP)[7] observations, scheduled arrival rate data, and NTML GDP records to calculate the hourly probabilities of a GDP being implemented. These probabilities were subsequently clustered using an Expectation Maximization (EM) algorithm[8] to identify similar weather impacted days at EWR and ORD. Finally, the 2012 GDP, GS and MIT records from the NTML database were analyzed to determine the frequency at which these restrictions were used under each of the similar types of days.

The overall modeling approach is presented in Section II. The LAMP, NTML and Aviation System Performance Metrics (ASPM) data used in the clustering and classification models, as well as the expression for calculating the probabilities of a GDP occurring at EWR and ORD, are presented in Section III. The similar types of days identified by the clustering algorithm are presented in Section IV along with an analysis focused on exploring the type and frequency of TMIs implemented on each of these days. Finally, a summary is presented in Section V.

II.   Modeling Methodology

A high-level architecture that illustrates how historical weather, air traffic and TMI databases can be used to support traffic flow management planning, which will subsequently be referred to as the "Learning Automation Architecture," is presented in Section II.A. A key component of this architecture is the identification of time periods (e.g., hours, days, etc.) in the NAS with similar weather and traffic conditions. An approach for identifying these similarities from the airport-perspective is described in Section II.B. This approach relies on the clustering of hourly probabilities of a GDP occurring at a specific airport. The approach for calculating these probabilities at EWR and ORD is presented in Section II.C. Finally, the EM clustering algorithm that was adopted to identify similar days at EWR and ORD is described in Section II.D.

A. Learning Automation Architecture

A high-level overview of how the data resident in the NTML archives can be used to support traffic flow management planning is illustrated in Fig. 1. Starting with the box labeled [1] in Fig. 1, the daily, historical NTML archives can be aggregated to define a daily national plan (see box labeled [3] in Fig. 1). In parallel, clustering and classification algorithms can be applied to historical weather archives (see box [2]) to identify similar historical weather days in the National Airspace System (NAS) (see box [4]). The historical national plans can subsequently be evaluated and ranked using performance metrics[9] (see box [5]). This will eventually allow traffic flow managers to identify national plans that were designed to minimize or maximize a specific objective function, such as equity or throughput, as is illustrated in the box labeled [8] in Fig. 1. The outcomes of boxes [1] through [5] can subsequently be used to inform day-of-operations decision making, as illustrated by boxes [6]-[9] in Fig. 1. Starting with the current weather observations and forecasts (see box [6]), days with similar weather can be identified (see box [7]) and subsequently the national plans along with the accompanying performance metrics can be made available to TMCs (see box [8]). Although somewhat tangential to the approach outlined in boxes [1]-[9] in Fig. 1, it is worth noting that the historical NTML data (see box [1]) and historical weather archives (see box [2]) can be used to directly predict whether a TMI is required (see box [10], as will be briefly discussed below.
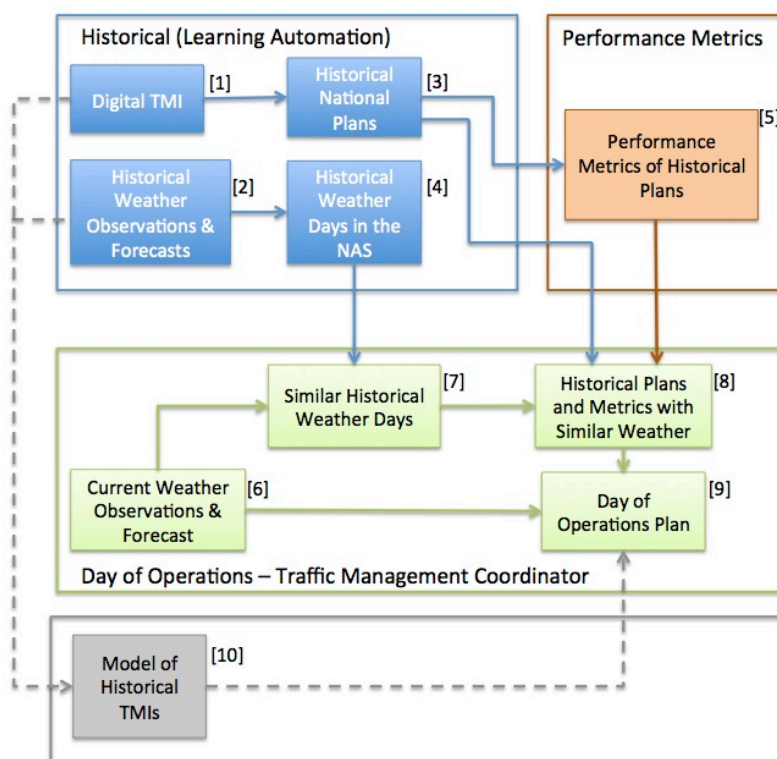

Figure 1. Overall approach to assisting day-of-operations planning with historical weather and air traffic data.

An overview of studies that have appeared in the literature that support the overall approach outlined in Fig. 1 follows. The works of Mukherjee et al.[10], Klein[11], Tien et al.[12] and Grabbe et al.[3] have all focused on identifying similar days in the NAS (see box [4] in Fig. 1). Mukherjee et al.[10] employed Ward's minimum-variance method[12] to cluster daily en route WITI[4] values to identify 21 national-level clusters and days within each cluster. Tien et al.[11] proposed a method for clustering weather-impacted scenarios based on an ensemble weather forecast product. Mukherjee et al.[10] presented the results of an initial analysis to better understand the TMIs used on similar weather days (see box [8] in Fig. 1), but this analysis focused exclusively on reroutes and en route weather constraints so additional research is needed in this area. Recent research focused on performance metric related work (see box [5] in Fig. 1) includes the work of Liu and Hansen.[13] Wanke and Taylor[14] have been developing a strategic traffic flow management decision-support system that is referred to as Flow Contingency Management (FCM), which can

directly support boxes [6]-[9]. It should be noted that the current work could provide historically similar airport clusters to FCM, as illustrated by the arrow leading from box [4] to [7] in Fig. 1.

Finally, studies by Wolfe and Rios[15], Smith et al.[16] Bloem et al.[17] and Wang and Grabbe[18] have all developed data mining models from historical data to directly predict whether or not a particular TMI is required on a given day (see box [10] in Fig. 1). Wolfe and Rios[15] utilized seven models from the information retrieval and machine learning community to determine if a Ground Delay Program was required at an airport given historical weather, traffic and operational actions. Similarly, Smith et al.[16] applied a support vector machine (SVM) to historical archives of Terminal Aerodrome Forecast (TAF) and Ground Delay Program (GDP) records to develop a model that could be used to predict GDPs at four Operational Evolution Partnership (OEP) airports. Bloem et al.[17] prototyped and evaluated four machine learning algorithms to estimate MIT restrictions for Newark Liberty International Airport arrivals. Finally, Wang and Grabbe[18] used an ensemble of decision trees with bootstrap aggregation to train a supervised binary classification model. This model was subsequently used to predict the need for arrival MIT restrictions at Hartsfield-Jackson Atlanta International Airport.

B. Modeling Approach

A diagram illustrating the overall approach that was adopted to identify clusters of hours with similar GDP probabilities at EWR and ORD is presented in Fig. 2. In the context of Fig. 1, the boxes in Fig. 2 represent a specific instantiation of the tasks associated with the "learning automation" boxes (i.e., the blue boxes) in Fig. 1. Boxes [1] through [4] in Fig. 2 depict the historical weather, air traffic, and TMI data sources that were leveraged in this study. The first stage of the modeling approach, which is depicted by box [5] in Fig. 2, entailed calculating the probability that a GDP occurred at a given hour at a particular airport. Details regarding these calculations are provided in Section II.A and III.B. Since GDPs are relatively infrequent events at many airports in the U.S., GDPs and GSs were treated interchangeably to improve the classification results; however, this is not a hard requirement. The motivation for calculating these probabilities instead of directly using the raw weather and air traffic data records was to improve the likelihood that the clusters identified would be operationally relevant (e.g., ones for which a TMC would have implemented a TMI). The calculation of these probabilities is discussed in Section II.B, and the resulting probabilities were subsequently clustered using an EM algorithm (see box [6] in Fig. 2), which is described in Section III.C. Lastly, the TMIs implemented in response to each type of cluster were analyzed, as indicated by box [7] in Fig. 2. For this analysis, GDPs and GSs were considered along with any MIT restriction that originated at either EWR or ORD (see box [4] in Fig. 2)
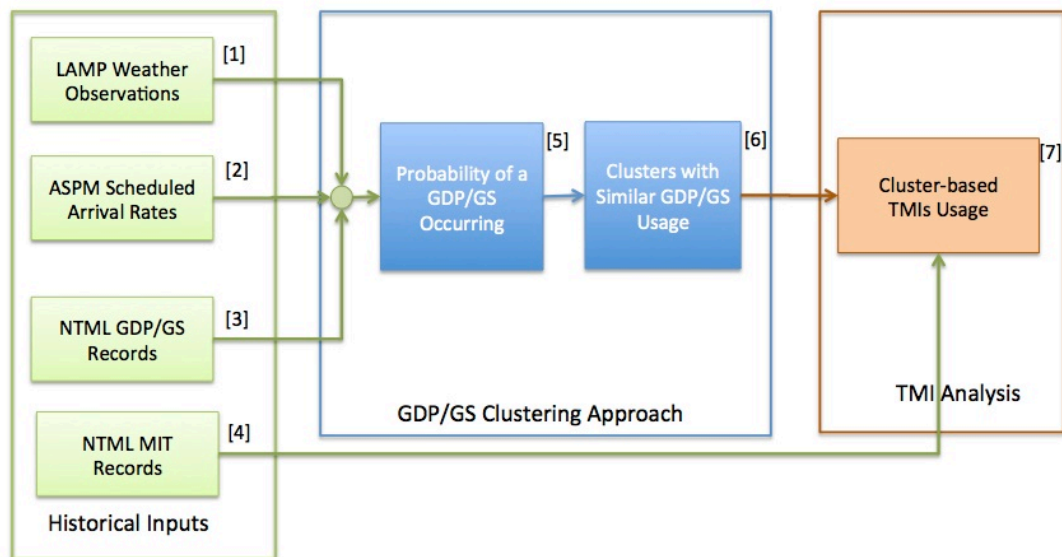

Figure 2. Modeling methodology for identifying similar weather impacted airport days.

C. Hourly Attribute Calculation

4

The set of hourly attributes (or features) that were used to cluster similar weather and arrival traffic conditions at EWR and ORD consisted of a set of probabilities that a GDP either was or was not implemented at each airport. These probabilities were calculated using a data mining technique known as a Logistic Model Tree (LMT).[19] For reference, a LMT is an approach that combines two popular supervised learning techniques, decision trees induction and logistic regression. To facilitate the discussion of the LMT approach, a brief overview of both decision trees and logistic regression follows.

1. Decision Tree Induction

The basic goal of decision tree induction is to divide a set of observations, such as meteorological and air traffic data for example, into regions with mostly similar class labels (e.g., GDP or NoGDP). The typical decision tree algorithm employees a top-down, greedy heuristic approach that recursively partitions the observations into two or more subgroups by setting a threshold on a selected attribute (e.g., wind speed > 20 kts). A wide range of techniques for selecting the attribute on which to partition the data and the corresponding threshold value have been proposed in the literature,[20] but most techniques seek to maximize a measure referred to as the "information gain" when making this decision.[20] After a decision tree is constructed in this top-down, recursive fashion, many of the branches in the tree will reflect outliers in the observations. Tree pruning[18] is a common technique to remove these outliers from the resulting decision tree.

The appeal of decision tree induction is that the algorithms are generally computationally efficient, and the results are easily interpreted. A path in a decision tree ultimately consists of a set of rules that indicate how to classify a set of observations. For more information about decision tree induction, the interested reader is referred to Reference 19.

2. Logistic Regression

The logistic regression model assumes that the probability of a GDP occurring, $\Pr(G = GDP|X = x)$, and the corresponding probability of no GDP occurring, $\Pr(G = NoGDP|X = x)$, can be calculated from Eqs. (1) and (2):

$$\Pr(G = GDP|X = x) = \frac{e^{F_{GDP}(x)}}{e^{F_{GDP}(x)} + e^{F_{NoGDP}(x)}} \tag{1}$$

$$\Pr(G = NoGDP|X = x) = \frac{e^{F_{NoGDP}(x)}}{e^{F_{GDP}(x)} + e^{F_{NoGDP}(x)}} \tag{2}$$

Here x is the vector of attribute values associated with a given instance of data, and $F_{GDP}(x)$ and $F_{NoGDP}(x)$ are linear combinations of the attribute values and a weighting coefficient. Note that Eqs. (1) and (2) sum to one.

3. Logistic Model Tree

The logistic model tree is a supervised learning technique that employs a tree induction algorithm to model possible complex structures in the data and capture nonlinearities in the data, and the leaves of the trees are simple logistic regression models. Two of the primary reasons that LMTs were adopted for this study are (1) explicit class probability estimates are generated by the model for each leaf, as opposed to the simple class values that are produced by decision tree induction algorithms, and (2) relevant attributes at each leaf are naturally selected by the logistic regression model. Once the LMT model develops the overall decision tree, the probabilities of a GDP occurring at each leaf of the tree are given by Eqs. (1) and (2).

To illustrate the use of LMTs, consider a simple example where the probability of a GDP occurring at EWR in any given hour is only related to the wind speed, v, and ρ=(hourly scheduled arrival demand/hourly arrival capacity). Using the LMT implemented in Weka[20], the truncated LMT appearing in Fig. 3 can be generated. The root node in this sample tree divides the dataset based on the value of ρ (e.g., ρ <= 0.86 or ρ > 0.86). Following the left branches of the tree, Fig. 3 illustrates that the dataset is once again divided based on the value of ρ, where data elements with ρ <= 0.55 form the left branch of the tree, and values with ρ > 0.55 form the right branch of the tree. The expressions $F_{GDP}(x)$ and $F_{NoGDP}(x)$ associated with two of the leaves of the tree are illustrated.
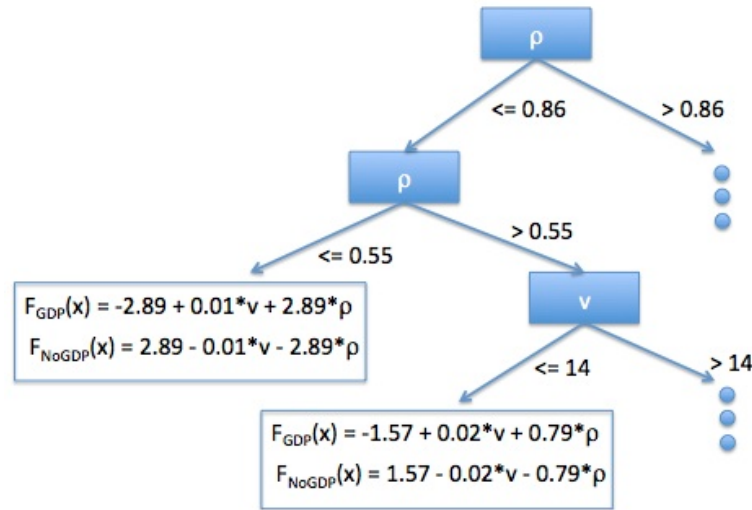
Figure 3. Truncated, sample Logistic Model Tree for EWR.

From the values of $F_{GDP}(x)$ and $F_{NoGDP}(x)$ appearing in Fig. 3, the probability of a GDP occurring, $\Pr(G = GDP|X = x)$, and the corresponding probability of no GDP occurring, $\Pr(G = NoGDP|X = x)$, can be calculated from Eqs. (1) and (2).

These probabilities subsequently served as the attributes in the clustering algorithm, which is described in Section II.C, to identify periods of time during which EWR and ORD operated under similar GDP likelihood conditions.

D. Similar Day Clustering Approach

Once the hourly probabilities associated with a GDP occurring or not occurring were calculated using the approach outlined in Section II.B, these values were subsequently clustered using the unsupervised Expectation Maximization (EM)[8] algorithm resident in the software package,[21] as illustrated by box [6] in Fig. 2. For reference the EM algorithm is an extension of the k-means algorithm that assigns each object to a cluster based on a probability of membership in each cluster. Validation of the clustering results was subsequently accomplished via ten-fold cross validation using the Weka software package. For this validation, the daily WITI factors were randomly partitioned into ten subsamples. Of the ten subsamples, a single sample was retained for validation while the remaining nine samples were used as training data for building the model. The cross validation process was subsequently repeated ten times with each of the ten samples used once as the validation data.

The hourly clustering results for EWR and ORD are presented in Section IV.

III.   Experimental Setup

This section provides an overview of the weather and air traffic data sources used, and provide a high-level overview of the LMTs developed for EWR and ORD that ultimately supplied the probabilities, which were used to develop the clustering results presented in Section IV.

A. Weather and Air Traffic Data

1. Localized Aviation MOS (Model Output Statistics) Program (LAMP)

The LAMP[7] weather observations were ideal for the purposes of this study because of (1) the availability of the data, (2) the compact format in which the observations and forecasts are stored, (3) the machine readability of the data, and (4) the long look-ahead forecasts (i.e., 24 or 25 hours) provided by the model. LAMP is developed by the National Weather Service (NWS), and provides forecasts at 1,692 stations across the continental United States, Alaska, Hawaii, Puerto Rico and the Caribbean Islands on an hourly basis. At any given station, the forecast is valid within a five statute mile radius from the center of the runway complex.

Although a total of 15 parameters are available with a LAMP forecast, the 12 weather parameters that are listed in the first 12 rows of Table 1 were used for classification purposes with the LMT. Noted that a solid horizontal line

is provided in the table to separate the LAMP parameters from the non-LAMP parameters. These parameters were selected to capture the major weather factors that are commonly associated with delays and TMIs at airport, such as wind speed, wind direction, ceilings, and probability of precipitation. Factors, such as, the temperature and dew point were not included as part of this initial study.

Table 1.   Weather and Traffic Attributes Used by the Logistic Model Trees.

Parameter Name	Description	Note
WDR	10-meter wind direction (tens of degrees)	-
WSP	10-meter wind speed (knots)	-
PPO	Probability of precipitation	-
PCO	"Yes" or "No" forecast of precipitation	Y, N, NA indicates variables not available
TP2	Probability of thunderstorms over a 2-hr period	-
TC2	"Yes" or "No" forecast of thunderstorms	Y or N
POZ	Conditional probability of freezing precipitation	Available Sept. 1 to May 31
POS	Conditional probability of snow occurring	Available Sept. 1 to May 31
TYP	Conditional precipitation type	Available Sept. 1 to May 31. S = snow, Z = freezing rain or drizzle, R = pure rain or drizzle or mixed with snow
CLD	Forecast category of sky cover	CL = Clear, FW = few clouds, SC = scattered, BK = broken clouds, OV = overcast
CIG	Categorical ceiling height	Eight ceiling categories defined
VIS	Categorical visibility	Seven visibility categories defined
ρ	ASPM Scheduled Arrival Demand / ASPM Arrival Capacity	
WITI	Center Weather Impacted Traffic Index	See Section III.A.3
GDP and GS Records	NTML/Digital TMI GDP and GS Records	Used to classify a hour as a GDP or no-GDP hour

2. Center WITI

In order to capture the en route impacts that convective weather can have on EWR and ORD arrivals, the Center-level WITI values for New York Center and Chicago Center, respectively, were calculated. Here the Center-level WITI is a proxy for the number of aircraft affected by convective weather in the en route environment, and is calculated at time *k* using Eq. (3).

$$WITI(k) = \sum_{i=1}^{m(k)} \sum_{j=1}^{n(k)} \delta_{ij} \tag{3}$$

American Institute of Aeronautics and Astronautics

Here $m(k)$ is the number of aircraft of interest in the Center of interest at time k, $n(k)$ is the number of weather polygons of interest in the Center at time k, and δ_{ij} is a Kronecker delta function. The delta function takes on a value of one when aircraft i is in weather cell j at time k, and is zero otherwise. For the purpose of this calculation, air traffic position data were obtained from the FAA's Aircraft Situation Display to Industry data feed, and convective weather polygons were defined as Corridor Integrated Weather System (CIWS)[22] polygons with vertically integrated liquid (VIL) levels of three or greater. This formulation of the WITI metric does differ from the original formulation presented in Ref. 9, but conceptually the two formulations are expected to yield similar results.

B. Logistic Model Tree

Using the 8,760 hourly weather and air traffic records from 2012, separate LMTs were developed for EWR and ORD where the classifiers associated with each hour specified whether or not a GDP had been used for the given hour. The rationale for undertaking this approach was to directly identify clusters of hours where GDPs were likely to occur. Weather and traffic conditions associated with the GDP-type clusters would likely be ones that would require some amount of intervention by TMCs in the operational environment. The complete tree for EWR is shown in Fig. 4.

The root node of the tree in Fig. 4 is associated with the attribute ρ, which is intuitive since GDPs are typically not required when the demand is far below the available capacity. Other important attributes associated with the EWR tree include the wind speed (wsp), the en route Center WITI (witi) and the probability of precipitation (ppo). An example of the expression for calculating $F_{1,GDP}(x)$ for the first leaf node in Fig. 4 is given by Eq. (4). It is worth noting that $F_{1,NoGDP}(x) = -F_{1,GDP}(x)$. In terms of the nomenclature, the index (1 in the case) appearing in the subscript of the leaf node is simply used to enumerate the leaf nodes.
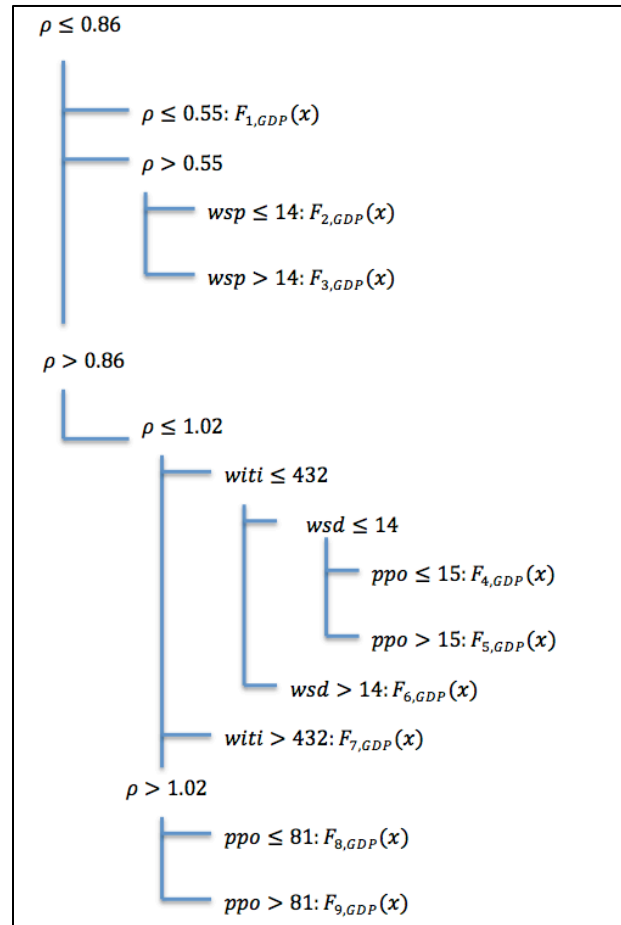

Figure 4. Complete Logistic Model Tree with GDP Classifiers for EWR.

American Institute of Aeronautics and Astronautics

By examining the sign and the magnitude of the weighting coefficients associated with the leaf nodes, the relative importance of each of the attributes associated with the leaf node can be ascertained. For example, by examining the expression associated with the first leaf node, $F_{1,GDP}(x)$), one finds that the likelihood of a GDP increases with increasing wind speeds (wsd), decreasing ceilings (cig), decreasing visibility (vis), increasing ρ, the presence of snow (i.e., [typ=S]) and broken clouds (i.e., [cld=BK]), as indicated by Eq. (4). Somewhat unintuitive, the likelihood of a GDP associated with the first leaf node is found to increase when the categorical forecast of precipitation over the next hour is "no" (i.e., [pco=N]) and there are clear skies (i.e., [cld=CL]). This inconsistency between the expected meteorological behaviors and the resulting GDP outcome is likely due to a lack of training data at this node, or possibly uncertainty in the forecast that was used when planning the GDPs. In future studies these results will be revised to include several year's worth of historical training data in an effort to eliminate some of these inconsistencies. The resulting probabilities of a GDP occurring or not are calculated from Eqs. (1) and (2).

$$F_{1,GDP}(x) = -2.12 + 0.06 * wsp - 0.1 * cig - 0.09 * vis + 2.84 * \rho + 0.16 * [typ = S] - 1.85 * [typ = Z] - 0.01 * [typ = R] - 0.07 * [typ = NA] - 0.05 * [pco = Y] + 0.08 * [pco = N] + 0.31 * [pco = NA] + 0.15 * [cld = CL] - 0.09 * [cld = FW] - 0.15 * [cld = SC] + 0.16 * [cld = BK] - 0.03 * [cld = OV] \quad (4)$$

For the ORD scenario with GDP classifiers, the resulting decision tree consists of a single node, and $F_{j,GDP}(x)$ is given by Eq. (5). As in the EWR scenario, $F_{1,NoGDP}(x) = -F_{1,GDP}(x)$.

As was done in the EWR case, the sign and magnitude of the weighting coefficients in Eq. (5) can be examined to determine the relative importance of each attribute when calculating the probability of a GDP at ORD. From Eq. (5), one finds that the probability of a GDP increases with increasing wind speeds (wsd), increasing probabilities of precipitation (ppo), decreasing ceilings (cig) and visibility (vis), increasing ρ, the presence of rain (i.e., [typ=R]), broken clouds ([i.e., cld=BK]) and overcast clouds ([i.e., cld=OV]). Using Eqs. (1), (2) and (5), the hourly probabilities of a GDP occurring at ORD can be calculated when keeping in mind that $F_{1,NoGDP}(x) = -F_{1,GDP}(x)$.

$$F_{1,GDP}(x) = -2.43 + 0.04 * wsp + 0.01 * ppo - 0.09 * cig - 0.14 * vis + 1.76 * \rho - 0.02 * poz + 0.15 * [typ = R] + 0.26 * [typ = N] + 0.36 * [pco = N] - 2.09 * [pco = NA] - 0.15 * [cld = CL] - 0.53 * [cld = FW] + 0.16 * [cld = BK] + 0.19 * [cld = OV] \quad (5)$$

In terms of the block diagram in Fig. 2, the results from this subsection are used to calculate the "Probability of a GDP/GS Occurring" (box [5]). These probabilities were subsequently clustered using the aforementioned EM clustering algorithm (see box [6] in Fig. 2), and these clustering results are presented in Section IV for EWR and ORD.

IV. Results

Using the 8,760 hourly GDP probability records for EWR and ORD from 2012 (see box [5] in Fig. 2), the EM clustering approach discussed in Section II.C was used to identify clusters of hours for which similar GDP probabilities were observed (see box [6] in Fig. 2). The GDP, GS and MIT usage statistics associated with each of these clusters was subsequently analyzed (see box [7] in Fig. 2). These clustering and TMI analysis results are presented in Section IV.A for ORD and Section IV.B for EWR.

A. ORD Clustering Results

The EM clustering algorithm identified a total of five clusters for the 2012 dataset. The number of members (i.e., hours) associated with each of the clusters and a brief description of each cluster is presented in Table 2. By far the largest cluster was Cluster 2, which contained approximately 57% of the hourly records. This cluster was associated with late night and early morning hours when ρ was generally low and the probabilities of a GDP occurring were low. Two additional clusters that are composed predominantly of low probability GDP events are Clusters 1 and 3. Both of these clusters are associated with fair to moderate daytime operations, and the majority of all hours associated with these cluster have GDP probabilities under 10%.

The two clusters associated with inclement weather at ORD are Clusters 0 and 4. Hours with reduced ceilings and elevated probabilities of precipitation characterize Cluster 0. GDP probabilities for the members associated

American Institute of Aeronautics and Astronautics

with this cluster range between 0 and 20%. In contrast, Cluster 4 is associated with the "bad weather" hours, which account for only 3.4% of all of the hourly events. The majority of members associated with this cluster have probabilities ranging from 30 to 100%.

Table 2. Description of 2012 ORD Clusters with GDP Classifiers.

Cluster Index	Number of Members	Description
0	660	Reduced ceilings and elevated probability of precipitation
1	1,833	Fair daytime weather hours
2	5,045	Early morning and nighttime operations
3	923	Moderate daytime weather hours
4	299	Bad weather – reduced ceilings and visibility, increased winds, increased probability of precipitation

To better understand the distribution of the probabilities of a GDP occurring for each Cluster, a stacked histogram of all hourly GDP probabilities that are organized by their cluster index is presented in Fig. 5. For presentation purposes, the GDP probabilities have been discretized into increments of 0.1, and it is worth noting that a logarithmic scale is being used on the vertical axis because of the large disparity in the number of members associated with each cluster. To better understand the data plotted in this figure, the leftmost bars indicate that 400 members of Cluster 0 (blue bars), 1,833 members of Cluster 1 (green bars), 5,045 member of Cluster 2 (salmon bars0 and 923 members of Cluster 3 (orange bars) have GDP probabilities less than 10%. The probabilities of GDP occurrence for Cluster 0 are nearly equally divided between members that have probabilities less than 10% (first blue bar in Fig. 5) and members whose probabilities are between 11% and 20% (second blue bar in Fig. 5). Although Cluster 4 is associated with typical "bad weather" events at ORD, Fig. 5 illustrates that GDP usage on these days varied widely (i.e., 11% to 100% chance of a GDP occurring).
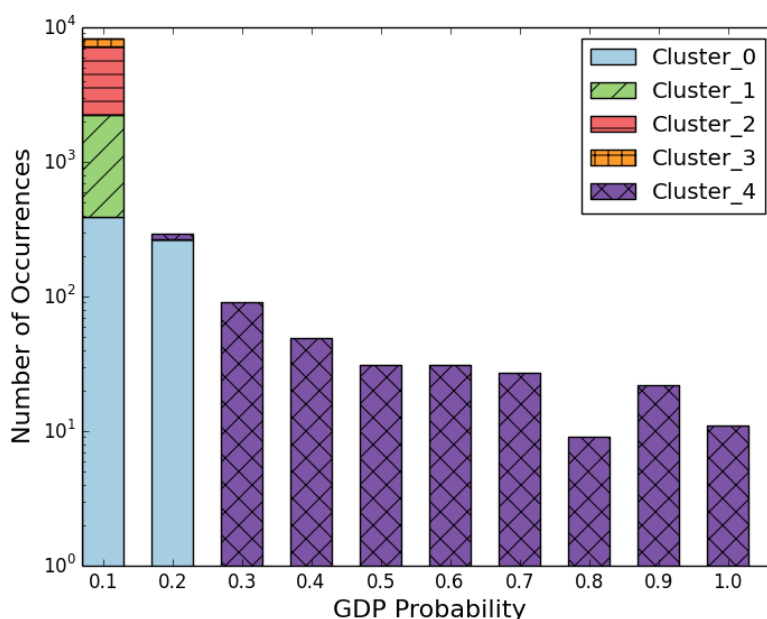

Figure 5. Stacked histogram of GDP hourly probabilities for the 2012 ORD Clusters.

In an effort to better understand the types of TMIs used for each of the five ORD clusters, each hour in 2012 was analyzed to determine if a GDP, GS or a MIT was impacting the airport. The results of the analysis are presented in Fig. 6. For reference, the size and color of the circles in Fig. 6 are both used to represent the relative frequency that

American Institute of Aeronautics and Astronautics

each TMI was used for each cluster. Although the size and color provide redundant information (e.g. a big circle or a red circle both convey a large number of data points), the color is included to aid in distinguishing between the different circle sizes, which becomes more important for the EWR results presented in the next subsection. For reference, a legend is also provided in the figure, which indicates the number of hours that a particular TMI was used for a particular cluster. For example, GDPs were issued for 153 of the hourly records (see the large red circle in Fig. 6) associated with Cluster 4, which is over 50% of the hourly records associated with this cluster. Recall from the discussion of Table 2 that Cluster 4 was the cluster associated with "bad weather" events at EWR. Cluster 0 was also associated with inclement weather at ORD, and an increased use of GDPs to mitigate these weather impacts is observed. Clusters 1, 2 and 3 are associated with generally good weather conditions, and the overall use of GDPs and GSs is observed to decrease for these clusters. Interestingly, there is a sharp increase in MIT usage for Clusters 1 and 2, which is likely required to control the increased arrival flows for the hours associated with these clusters.

Once the clustering results are extended to consider weather forecasts, the result could ultimately be used by traffic flow managers to identify historically similar air traffic and weather conditions. The TMIs issued under these historically similar conditions could be used to help inform day of operations planning by the traffic flow managers.
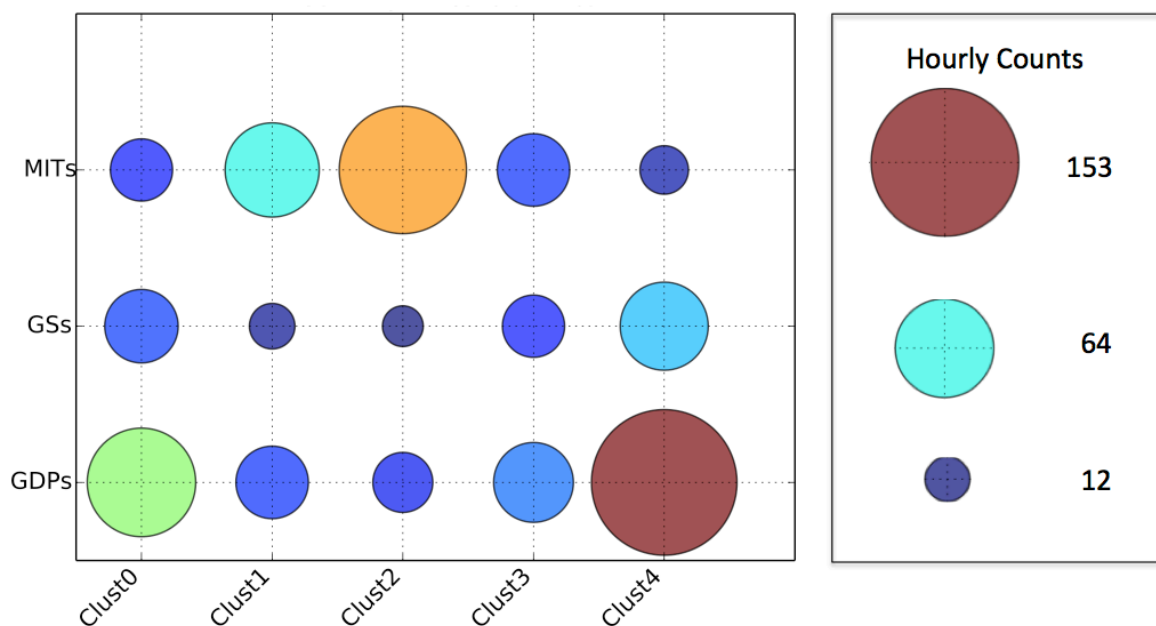

Figure 6. Mapping between TMIs and 2012 ORD Clusters with GDP Classifiers.

B. EWR Clustering Results

As compared to ORD, EWR experiences GDPs for a wider range of air traffic and weather conditions, and the corresponding set of clusters generated by the EM clustering algorithm is much richer. A total of 13 clusters were generated for EWR, and a brief description of these clusters is provided in Table 3.

The cluster with the most members is Cluster 2, which is characterized by nighttime operations where the traffic levels are reduced and the corresponding probability of a GDP is also low. This cluster contains roughly 26% of the hourly data records from 2012. Clusters 6 and 11 are both associated with morning operations where the arrival rates are increasing, but again there is typically insufficient demand to warrant extensive GDPs. Approximately 23% of the hourly observations reside within these "morning" clusters. Overall, roughly 49% of all hourly records were associated with either nighttime or early morning operations. Clusters 0, 1, 5, 8, 10 and 12 all have similar weather and traffic conditions where increased probabilities of snow and freezing precipitation are possible. Although similar weather and traffic conditions exist, there is observed to be a high degree of variability in the use of GDPs under these conditions, which will be discussed below. This could be due to differences in operator preferences, or it may indicated that additional features or attributes are being considered in the operational

American Institute of Aeronautics and Astronautics

environment which are not captured in this initial study. The weather attributes associated with Cluster 9 are similar to these clusters, however the probability of snow is considerably lower. Cluster 3 is associated with increased hourly Center WITI values and increased probabilities of precipitations. Members of this cluster are relatively rare and account for only 2% of all hourly records for 2012. Cluster 4 is associated with the afternoon arrival rush, where the arrival count is high and an elevated chance of precipitation also exists. Again, only about 2% of the hourly records are associated with this cluster. Finally, Cluster 7 is a classic "bad weather" scenario where the Center WITI is high, the probability of precipitation is high, and the ceilings and visibility are reduced. Recall from the introduction that EWR GDPs are typically associated with frontal passages that can lead to increased winds, low ceilings and poor visibility. For reference, clouds and weather normally concentrate near fronts, which are caused by frontal lifting. Fronts are regions of wind shear that can be identified by a wind direction shift. Low level wind shear follows a cold front and precedes the passage of a warm front. Low level wind shear lasts about one to three hours after a cold front and up to six hours before a warm front. Wind shear with a warm front lasts longer and frequently occurs with low ceilings and poor visibility.[24] Less than 2% of the hourly weather and traffic records reside within this cluster.

Table 3. Description of 2012 EWR Clusters with GDP Classifiers.

Cluster Index	Number of Members	Description
0	213	Increased probability of snow and freezing precipitation with slightly reduced ceilings
1	840	Increased probability of snow and freezing precipitation with reduced ceilings (very similar to Cluster 0)
2	2,256	Night time operations
3	196	Increased Center WITI and increased probability of snow and freezing precipitation
4	210	Hours associated with the afternoon arrival rush, high arrival demand hours with an elevated probability of precipitations
5	509	Increased probability of snow and freezing precipitation with slightly reduced ceilings (similar to Cluster 0)
6	1,078	Morning operations with reduced arrival counts and elevated probability of rain and snow
7	165	High Center WITI, increased probability of precipitation, reduced ceilings and visibility, and increased winds with frontal passage – "Bad Weather"
8	329	Increased probability of snow and freezing precipitation with slightly reduced ceilings (similar to Cluster 0)
9	942	Elevated probability of freezing precipitation
10	642	Increased probability of snow and freezing precipitation with slightly reduced ceilings (similar to Cluster 0)
11	923	Morning operations with increased probability of snow and freezing precipitation
12	457	Increased probability of snow and freezing precipitation with slightly reduced ceilings (similar to Cluster 0)

To better understand the distribution of the probabilities of a GDP occurring for each Cluster, a stacked histogram of all hourly GDP probabilities that are organized by their cluster index is presented in Fig. 7. For presentation purposes, the GDP probabilities have been discretized into increments of 0.1, and it is worth noting that a logarithmic scale is being used on the vertical axis because of the large disparity in the number of members associated with each cluster. To better understand the data plotted in this figure, the rightmost bar (tan with small circles) has a height of 165, which indicates that all 165 members of Cluster 7 have GDP probabilities greater than or equal to 90%. The figure clearly illustrates that all members of Clusters 2 (leftmost, green bar with black horizontal strips), which are associated with nighttime operations, have GDP probabilities of less than 10%. All 923

members of Cluster 11, which are associated with early morning operations, also have GDP probabilities of less than 10%. Cluster 6 (red bars with stars), which is also associated with morning operations, has GDP probabilities less than 20% because of the increased likelihood of rain and snow. The "bad weather" cluster is Cluster 7 (see rightmost bar in Fig. 7), and all members of this cluster have GDP probabilities of 90% or greater, which indicates that the decision to implement a GDP when the weather is very bad is very consistent. Some of the most interesting clusters identified by the EM algorithm are Clusters 0, 1, 5, 8, 10 and 12. Although the weather and traffic conditions associated with these clusters are similar, the probabilities of GDP occurrence vary widely across these six clusters. For example, most members of Cluster 0 have GDP probabilities greater than 70%, while the members of Cluster 1 have probabilities of GDP occurrence between 11% and 30%. Although there is a lot of decision inconsistency in whether a GDP will or will not be used under the weather and traffic conditions associated with these clusters, the algorithm has grouped together hours for which similar decisions were historically made.
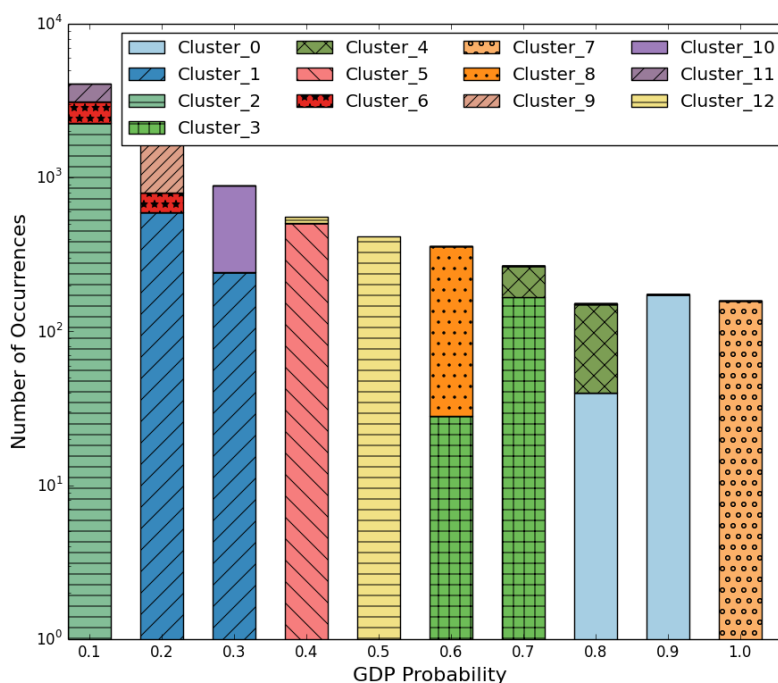

Figure 7. Stacked histogram of GDP hourly probabilities for the 2012 EWR Clusters.

As was done for the ORD scenario, the hours associated with the 12 EWR clusters were analyzed to determine if a GDP, GS or a MIT was impacting the airport, and the results are presented in Fig. 8. For reference, the size and color of the circles in Fig. 8 are used to represent the relative frequency that each TMI was used for each cluster. Although the size and color provide redundant information (e.g. a big circle or a red circle both convey a large number of data points), the color is included to aid in distinguishing between the different circle sizes. For reference, a legend is also provided in the figure, which indicates the number of hours that a particular TMI was used for a particular cluster. For example, GDPs were issued for 190 of the hourly records (see the large red circle in Fig. 8) associated with Cluster 12, which is over 40% of the hourly records associated with this cluster. Cluster 2, which is associated with nighttime operations at EWR, is observed to have only a few instances during which GSs were used and no MITs or GDPs. GDP usage for the early morning operation clusters (Clusters 6 and 11) are also observed to have relatively low numbers of GDPs, but MIT usage, especially for Cluster 6, is found to be on the increase. Presumably this increase is necessary to cope with the increasing arrival demand associated with the morning arrival rush. Cluster 11, which is associated with early morning operations, also had relatively few TMIs. Cluster 7, which is a "bad weather" cluster, has a relatively large number of GDPs, and a limited number of MITs and GSs. Clusters 0, 1, 5, 8, 10 and 12 were the six clusters with similar weather and traffic conditions, and all the clusters have relatively high numbers of GDPs but the MIT usage is found to very markedly across the clusters.

This of course is somewhat to be expected since the clustering did not take into account MIT usage, but this is an element worth exploring in future studies.
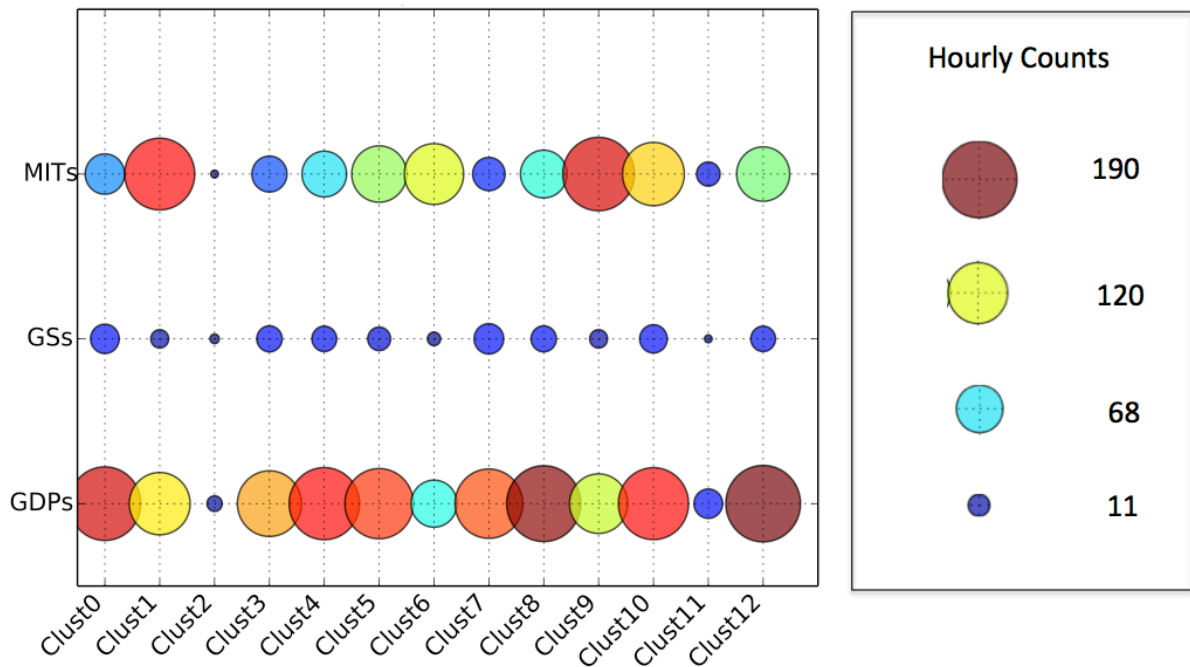

Figure 8. Mapping between TMIs and 2012 EWR Clusters with GDP Classifiers.

Based on the two airports analyzed in this section, it does appear as if the LMT classification approach and the accompanying EM clustering algorithm do lead to operationally relevant clusters. In cases where similar weather and traffic conditions lead to different decision making approaches, the clustering algorithm naturally creates separate clusters to reflect the variations in the operational decision making process. In future studies, this work should be extended to consider both weather forecasts at the airport and in the vicinity of the airport, in an effort to better understand the relationship between the forecasted weather conditions and the TMIs that were operationally implemented.

V.   Conclusions

Using a two-step process, clusters of hours with similar probabilities of Ground Delay Program usage were identified for Chicago O'Hare International Airport and Newark Liberty International Airport. In the first step, historical weather, scheduled arrival rates and traffic flow management logs from 2012 were used to estimate the hourly probability of a Ground Delay Program occurring at both Chicago O'Hare International Airport and Newark Liberty International Airport. These hourly probabilities were subsequently clustered in the second-step to identify collections of hours in which the likelihood of a Ground Delay Program was similar. The rational for clustering the Ground Delay Program probabilities as opposed to directly clustering the raw weather and air traffic attributes was to identify operationally relevant clusters where, for example, a traffic management coordinator may have been required to implement a traffic management initiative.

For Chicago, a total of five clusters were identified, while thirteen clusters were identified for Newark. This large difference reflects the increased diversity in air traffic and weather conditions that result in Ground Delay Programs being implemented at Newark, as compared to Chicago. For both airports, clusters associated with nighttime and early morning operations were clearly identifiable, and the probabilities associated with Ground Delay Programs occurring for these clusters was very low, as expected. These were by far the largest clusters for both airports and they contained between 49% of the hourly records for EWR and 57% of all the hourly records for ORD.

Additional clusters representing typical "bad weather" days were also identified where the probabilities of a Ground Delay Program occurring were high, but these were rare and accounted for less than 3% of all the hourly events. Somewhat surprising was the identification of six clusters for Newark where the weather and traffic conditions were similar, but the probabilities for a Ground Delay Program occurring were found to vary significantly. Finally, the frequencies of occurrence of Ground Delay Programs, Ground Stops and Miles-in-Trail restrictions implemented for each of the clusters were analyzed. Based on the results, Ground Delay Program usage and to a lesser extent Ground Stop usage was found to increase as the probability of a Ground Delay Program increased. The same trend was not observed for Miles-in-Trail restrictions, which is to be expected, since Miles-in-Trail usage was not considered in the clustering algorithm.

Finally, the results demonstrate that it is feasible to use historical weather and air traffic archives to provide guidance on the types of traffic management restrictions to implement in response to the weather and traffic conditions impacting an airport.

References

[1]Federal Aviation Administration, *Facility Operation and Administration*, Chap. 17-5, U.S. Department of Transportation, 2010.

[2]Klein, A., "Day's Weather in the NAS: Visualization of Impact, Quantification and Comparative Analysis," *Proceedings of the Integrated Communications Navigation and Surveillance (ICNS) Conference*, 2006.

[3]Grabbe, S., Mukherjee, A., and Sridhar, B., "Similar Days in the NAS: an Airport Perspective," *AIAA Aviation Technology, Integration, and Operations (ATIO) Conference*, Los Angeles, CA, Aug. 2013.

[4]Callaham, M. B., DeArmon, J. S., Cooper, A., Goodfriend, J. H., Moch-Mooney, D., and Solomos, G., "Assessing NAS Performance: Normalizing for the Effects of Weather," *4th USA/Europe Air Traffic Management R&D Seminar*, 2001.

[5]Han, J. and Kamber, M., Data Mining Concepts and Techniques, Morgan Kaufman Publishers, San Francisco, CA, 2001.

[6]Rios, J., "Aggregate Statistics of National Traffic Management Initiatives," *AIAA Aviation Technology, Integration, and Operations (ATIO) Conference*, AIAA-2010-9382, Fort Worth, TX, Sept. 2010.

[7]Ghirardelli, J. E., "An Overview of the Redeveloped Localized Aviation MOS Program (LAMP) for Short-Range Forecasting," American Meteorological Annual Meeting, 2005.

[8]Do, C. B. and Batzoglou, S., "What is the expectation maximization algorithm?," *Nature Biotechnology*, Vol. 26, No. 8, Aug. 2008.

[9]Bradford, S., Knorr, D., and Liang, D., "Performance measures for future architecture," *3rd USA/Europe Air Traffic Management R&D Seminar*, 2000.

[10]Mukherjee, A., Grabbe, S., and Sridhar, B., "Classification of Days Using Weather Impacted Traffic in the National Airspace System," *AIAA Aviation Technology, Integration, and Operations (ATIO) Conference*, Los Angeles, CA, Aug. 2013.

[11]Tien, S. L., Taylor, C. and Wanke, C., "Identifying Representative Weather Scenarios for Flow Contingency Management," *AIAA Aviation Technology, Integration, and Operations (ATIO) Conference*, Los Angeles, CA, Aug. 2013.

[12]Ward, J. H., "Hierarchical Grouping to Optimize an Objective Function," *Journal of American Statistical Association*, Vol. 58, No. 301, pp. 236-244, 1963.

[13]Liu Y. and Hansen M. "Ground Delay Program Decision-making using Multiple Criteria: A Single Airport Case," *10th USA/Europe Air Traffic Management R&D Seminar*, 2013.

[14]Wanke, C. and Taylor, C., "Exploring Design Trade-offs for Strategic Flow Planning," *AIAA Aviation Technology, Integration, and Operations (ATIO) Conference*, Los Angeles, CA, Aug. 2013.

[15]Wolfe, S. R. and Rios, J. L., "A Method for Using Historical Ground Delay Programs to Inform Day-of-Operations Programs," *AIAA Guidance, Navigation and Control Conference*, AIAA 2011-6528, Aug. 2011.

[16]Smith, D. A., Sherry, L., and Donohue, G., "Decision Support Tool for Predicting Aircraft Arrival Rates, Ground Delay Programs and Airport Delays from Weather Forecasts," *Proceedings of the International Conferences on Research in Air Transportation (ICRAT)*, 2008.

[17]Bloem, M., Hattaway, D., and Bambos, N., "Evaluation of Algorithms for a Miles-in-Trail Decision Support Tool," *Proceedings of the International Conference on Research in Air Transportation (ICRAT)*, 2012.

[18]Wang, Y. and Grabbe, S., "Modeling Weather Impact on Airport Arrival Miles-in-Trail Restrictions," *SAE Journal of Aerospace,* Vol. 6, No.1, pp. 247-259, 2013.

[19]Landwehr, N., Hall, M., and Frank, E., "Logistic Model Trees," Machine Learning, Vol. 95, No. 1-2, pp. 161-205, 2005.

[20]Rokach, L. and Maimon, O., *Data Mining with Decision Trees: Theory and Application (Series in Machine Perspective and Artificial Intelligence)*, World Scientific Publishing Company, 2008.

[21]Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H., "The WEKA Data Mining Software: An Update," SIGKDD Explorations, Vol. 11, No. 1, 2009.

[22]Evans, J. E. and Ducot, E. R., "Corridor Integrated Weather System," Lincoln Laboratory Journal, Vol. 16, No. 1, pp. 59-80, 2006.

[23]Wang, Y. and Kulkarni, D., "Modeling Weather Impact on Ground Delay Programs," *SAE Journal of Aerospace,* Vol. 4, No.2, pp. 1207-1215, 2011.

[24]Lester, P.F., *Aviation Weather*, 3rd edition, Jeppesen, Englewood, CO, 2007.