

# Similar Days in the NAS: an Airport Perspective

Shon Grabbe\* and Banavar Sridhar†  
NASA Ames Research Center, Moffett Field, CA, 94035-1000

Avijit Mukherjee‡  
University of California Santa Cruz, Moffett Field, CA, 94035-1000

On any given day, constraints in the National Airspace System, for instance weather, necessitate the implementation of Traffic Flow Management initiatives, such as Ground Delay Programs. The parameters associated with these initiatives, for example the location, scope, duration, etc., are typically left to human decision makers, who must rely on intuition, past experience, and weather and traffic forecasts. Although the decisions of these traffic flow specialists are recorded on a daily basis, few studies have attempted to apply data mining techniques to these archives in an attempt to identify patterns and past decisions that could ultimately be used to influence future decision-making. The goal of this study is to take a preliminary step towards informing future decision-making by proposing a technique for identifying similar days in the National Airspace System in terms of the Ground Delay Programs that were operationally implemented. Hence an airport perspective is being taken to identify these similar days, as opposed to considering possible airspace features. A modified k-means clustering algorithm is applied to all days in 2011, resulting in the identification of 18 clusters that represent unique combinations of Ground Delay Program that were historically implemented. A given day was described in terms of the presence or absence of 33 features that were a combination of Ground Delay Program locations and causes. By far the largest cluster that was identified consisted of 73 days on which low ceiling related Ground Delay Programs impacted San Francisco International Airport. In an attempt to verify the stated cause of the Ground Delay Programs, an Expectation Maximization clustering algorithm was applied to the 8,760 hourly Meteorological Aerodrome Reports, scheduled arrival rate and Ground Delay Program start and end time records for 2011. In general, clusters were identified that corroborated the stated causes of the Ground Delay Programs. However, these clusters often contained a significant number of members for which a Ground Delay Program did not occur. Findings from this initial study indicate that it is possible to identify similar days under which the National Airspace System operates, and clustering techniques appear to be promising methods for identifying the major causes of Ground Delay Programs.

## I. Introduction

SEVERE weather in the National Airspace System (NAS) can reduce both airport and airspace capacity. When this occurs, traffic flow managers rely on a collection of 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 any demand-and-capacity imbalances in the NAS. Although tools exist for implementing GDPs, AFPs and reroutes the parameters associated with these TMIs (e.g., start times, stop times, airport arrival rates, etc.) are often left to traffic flow specialists within the FAA and dispatchers from the Airline Operation Centers (AOCs) to determine. With limited “what-if” simulation capabilities, the human operators must

---

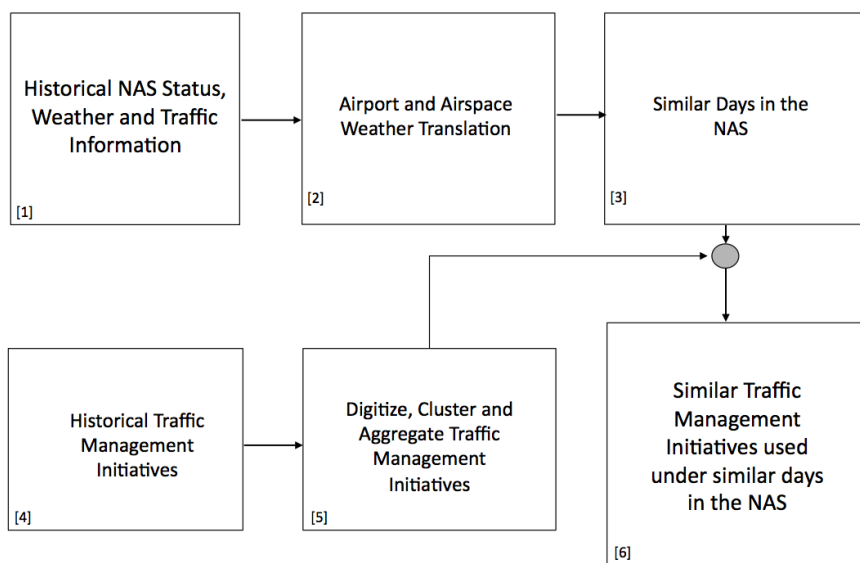
\* Research Scientist, Systems Modeling & Optimization Branch, Mail Stop 210-15, shon.grabbe@nasa.gov, Associate Fellow AIAA.

† Senior Scientist, Aviation Systems Program, Mail Stop 210-10, Fellow AIAA.

‡ Associate Project Scientist, University of California Santa Cruz, Mail Stop 210-8.

often rely on past experience and intuition when determining these parameters, even though past control actions have been digitally archived since the deployment of the National Traffic Management Log (NTML)<sup>1</sup> in the mid-2000's.

To determine if these historical archives can be used to support post operations analyses, and ultimately day-of-operations planning, a new line of research is being undertaken that seeks to leverage models from the machine learning and data mining communities. Research supporting this activity is being divided along two parallel paths as illustrated by the overall approach presented in Fig. 1. In the first path (see boxes labeled [1], [2] and [3] in Fig.1), historical NAS status information (e.g., runway configurations, runway construction, etc.), weather and traffic counts are being collected and archived (see box [1] in Fig. 1). The weather observations and forecasts are subsequently translated into airport and airspace impacts (see box [2] in Fig. 1). It is worth noting that a sizeable number of weather translation studies have already appeared in the literature (see for example, Ref. 2), and to the extent possible these models will be leveraged in support of this research effort. After translating the weather information, machine learning and data mining techniques will be used to identify similar days in the NAS. In parallel with this activity, archives of historical TMIs (see box [4] in Fig. 1) are being digitized<sup>3</sup>, clustered and classified (see box [5] in Fig. 1) to reduce the dimensionality of the archived data and to identify patterns and trends in the data. Lastly, this area of work will attempt to combine the research designed to identify similar days in the NAS (see box [3] in Fig. 1) with the clustered and classified TMIs (see box [5] in Fig. 1) to determine if similar TMIs are typically used on similar days in the NAS (see box [6] in Fig. 1). Associated with the work required for box [6] is the need to assess the performance (e.g., throughput, equity, efficiency, etc.) of the NAS on days under which similar constraints impacted the NAS, but a potentially differing sets of TMIs were used in response to the constraints. The results presented in this study are solely focused on contributing to the research associated with box [3] in Fig. 1. Follow-on studies are underway or ongoing to address the research associated with boxes [2], [5] and [6].



**Figure 1. Roadmap to similar traffic management initiatives under similar days activities.**

Past work supporting the activities outlined in Fig. 1 includes the study by Mukherjee et al.<sup>4</sup> that employed the Ward's minimum-variance method<sup>5</sup> to cluster daily en route Weather Impacted Traffic Index (WITI)<sup>6</sup> values to identify 21 national-level clusters and days within each cluster. Wolfe and Rios 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.<sup>7</sup> Similarly, Smith et al.<sup>8</sup> applied a support vector machine (SVM) to historical archives of Terminal Aerodrome Forecasts (TAFs) and Ground Delay Programs (GDPs) records to develop a model that could be used to predict GDPs at four OEP airports. Finally in Klein<sup>9</sup>, a "day-at-a-glance weather impact matrix" is presented for visualizing similarities of weather impacts both at the NAS-level, as well as, the airport level.

Building on the work of Klein and Mukherjee et al. an approach is proposed for identifying similar days in the National Airspace System (NAS) (see box [3] in Fig. 1) in terms of the frequency, distribution and cause of historical GDPs. This is an important first step as the next step includes identifying the actions taken on these days and their resultant NAS performance. To accomplish this, the cause and location of all GDPs implemented in 2011 were analyzed and subsequently clustered using a standard K-means clustering algorithm that was adapted to use the Jaccard index<sup>10</sup> to measure similarity between the cluster members. A total of 18 unique clusters and the associated days within the clusters were identified using this technique. The work was further extended by analyzing historical Meteorological Aerodrome Report (METAR), scheduled arrival rates and GDP start and end time records in an attempt to confirm the cause of the GDPs that were listed in the NTML using standard data mining clustering techniques.

The clustering models, clustering features and clustering stopping criteria are presented in Section II. The NTML GDP data and METAR data used in the clustering models is presented in Section III. Clustering results are presented in Section IV. Finally, a summary is presented in Section V.

## II. Modeling Methodology

Two complimentary clustering approaches are employed in this study to first identify similar days in the NAS from an airport perspective, and second to cluster weather and air traffic data to identify the traffic and meteorological conditions under which GDPs are implemented at select airports. The former of these clustering approaches, which is described in Section II.A, will result in national-level clusters where the features used for clustering are the NTML stated causes of the GDPs at select airports in the NAS (hence the “airport perspective” name). The later of these clustering approaches, which is described in Section II.B, is designed to determine if the GDP stated causes that are used in Section II.A are substantiated by historical weather and traffic observations. For example, if a GDP for winds is implemented at Newark, do the weather observations indicate strong winds were impacting the airport?

### A. National-level Clusters

To identify similar days in the NAS in terms of the location and cause of the historical GDPs, a simple k-means clustering algorithm was employed. Briefly, the algorithm attempts to partition a set of  $N$  objects (days in our case) into  $K$  clusters, such that the intra-cluster similarity is high while the inter-cluster similarity is small.<sup>10</sup> Here the cluster similarity is measured with respect to the mean value of the objects in the cluster. For our study, the features (GDP cause and location) associated with each of the  $N$  objects (days) were Boolean variables. A value of 1 indicated that a GDP impacted a particular airport on a particular day for a specific cause, and a value of 0 indicated that there was no corresponding GDP at the airport. As an example, “SFO+LowCeilings” is a feature that was considered in the clustering model to indicate that a GDP due to low ceilings was implemented at San Francisco International Airport (SFO) on a given day. A more extensive discussion of the 33 features used in the clustering model is presented in Section III.

Although the typical k-means clustering algorithm utilizes the Euclidean distance, this measure is ill suited for measuring distances between Boolean variables. Instead, the Jaccard coefficient<sup>10</sup> was used as the distance metric. Given days  $A, B \in N$  each with  $n$  binary features, the Jaccard coefficient is defined as

$$J_{AB} = \frac{M_{01} + M_{10}}{M_{01} + M_{10} + M_{11}} \quad (1)$$

Here  $M_{01}$  is the total number of instances where the features of  $A$  is 0 and the feature of  $B$  is 1;  $M_{10}$  is the total number of instances where the feature of  $A$  is 1 and the feature of  $B$  is 0 and finally  $M_{11}$  is the number of instances where the feature of  $A$  is 1 and the feature of  $B$  is 1. Clearly,  $J_{AB} \in [0,1]$  and is maximal when the features associated with the two days are completely dissimilar (e.g.,  $M_{01} + M_{10} = 1$ ), and is minimal when the features associated with the two days are entirely similar (e.g.  $M_{10} + M_{01} = 0$  and  $M_{11} = n$ ). When calculating this metric, situations in which neither day possess a feature are excluded from the calculation. For computational purposes, if neither day possesses any of the features then  $J_{AB} = 1$  (e.g., the two days are considered to be dissimilar).

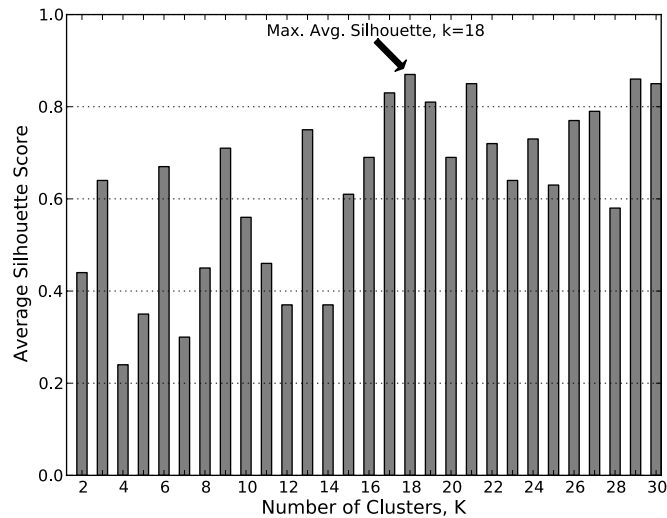
One of the key parameters that must be provided to the k-means clustering algorithm is the number of clusters,  $K$ . To determine a reasonable value of  $K$  for our application, the algorithm was iteratively run with  $K \in [2,30]$ , and the “best” value of  $K$  was selected that maximized the average Silhouette score<sup>10</sup> over all days. Here the Silhouette score lies within  $[-1,1]$  and is a measure of how well a given object (day) fits within the other days in a given cluster. A value of 1 for the Silhouette score indicates that the object is within an appropriate cluster, a value of -1 indicates

that the object would have been more appropriately placed in a different cluster, while a value of 0 indicates that the object is on the border of belonging to two different clusters. The Silhouette score for object (day)  $i$ , can be written as

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (2)$$

Here  $a_i$  is the average distance of object  $i$  to all other objects in the cluster that  $i$  has been assigned to. Similarly  $b_i$  is the lowest average distance between object  $i$  and the members of another cluster that  $i$  does not belong to.

The average silhouette score as a function of the number of clusters,  $K$ , is shown in Fig. 2. The maximum average silhouette score was 0.87, which occurred when  $k=18$ . This value of  $K$  was subsequently selected as the “best” value for generating the national-level clusters. In generating this figure, the data for all days in 2011 was repeatedly clustered into 2 through 30 different clusters and the average silhouette score across all days for each of the different number of clusters is plotted in Fig. 2. For example, in order to generate the left-most bar in Fig. 2, all data from 2011 was clustered into two different clusters and the average silhouette score for all of the days in 2011, which turned out to be 0.42, is being plotted in this figure.



**Figure 2. Average silhouette score as a function of the number of clusters.**

The results of the national-level GDP clustering study are presented in Section IV.

### B. Airport-level Clusters

Although the NTML lists the causes (e.g., low ceilings) of the GDPs, the weather observations and forecasts in and around the airport must ultimately be translated into airport-level impacts in order for the data mining approaches described in this study to be applicable for day-of-operations planning. Much of the airport-level weather translation modeling work appearing in the literature (see for example, Refs. 11 and 12) has been focused on calculating the weather impacted airport arrival and departure rates, and far less research has focused on determining if a GDP is required at a specific airport (see for example, Refs. 7 and 8).

Since developing a new weather translation algorithm to determine if a GDP is required at a given airport is beyond the scope of this study, a simpler approach was adopted where clustering models were used in an attempt to verify the stated causes of the GDPs. For example, if “low ceilings” is the cause of a GDP listed in NTML, were the ceilings in fact low during the GDP? Of course, there are many nuances to this problem, which will be explored in more detail in future studies. For example, if “low ceilings” are the stated problem, are the low ceilings directly at the airport or further away from the airport where controllers are attempting to sequence and space the arrivals for final approach? For this initial study, only the weather observations at the airport (e.g., METAR data) was considered, and more refined airport specific models will be the subject of future investigations.

All airport-level clustering was accomplished using the Expectation Maximization (EM)<sup>13</sup> clustering algorithm resident in the Weka software package.<sup>14</sup> 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 10-fold cross validation using the Weka software package. For this validation, the 8,760 hourly METAR, scheduled arrival rate and GDP start/end time data samples for each of the analyzed airports from 2011 were randomly partitioned into 10 subsamples. Of the 10 subsamples, a single sample was retained for validation while the remaining 9 samples were used as training data for building the model. The cross validation process was subsequently repeated 10 times with each of the 10 samples used once as the validation data.

As will be described in Section III.B, eight traffic and weather features were available for the airport-level clustering. However, not all of the features were expected to be equally relevant at each of the airports considered in this study. For this reason, a feature relevance analysis was performed for each of the selected airports prior to applying the EM clustering algorithm using the correlation-based feature selection algorithm (CfsSubsetEval) in the Weka software package.<sup>13</sup> Briefly, this algorithm attempts to determine the merit of a subset of features by considering the individual predictive ability of each feature (e.g., how well does a given feature determine if a GDP is impacting an airport or not). Subsets of features that are highly correlated with either the presence or absence of the GDP at a particular airport for a particular cause are selected by the algorithm.

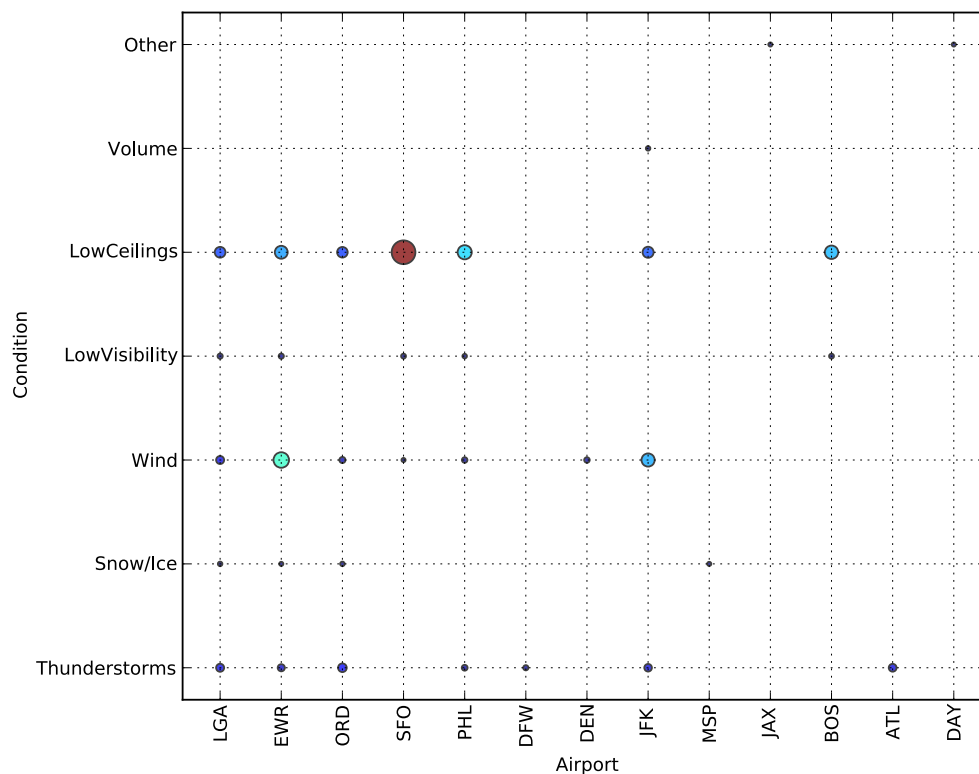
The results of the attribute relevance analysis and the airport-level clustering results are presented in Section IV.B for SFO, Newark Liberty International Airport (EWR) and Hartsfield-Jackson Atlanta International Airport (ATL).

### III. Experimental Setup

#### A. National-level Clusters

As previously mentioned, the location and stated cause for the GDPs in 2011 served as the basis for the national-level airport clustering study. For 2011, there were a total of 957 GDP records. However, a large number of these events occurred only once throughout the year. For example, a security related GDP was implemented at JFK on May 5, 2011. To improve the clustering performance, only GDPs that occurred at a particular airport at least five times throughout the year were used in the clustering process. The value of five was selected solely based on engineering judgment, and in future studies a more systematic approach for selecting this value is warranted. As a result of this filtering, a total of 788 GDP events were considered.

The location and cause of these GDPs are presented in Figure 3. For reference, the size and color of the circles in Figure 3 are used to represent the relative frequency of each GDP. Higher frequency events are indicated by bigger circles that are represented by “warmer” colors (e.g., red), while lower frequency events are indicated by smaller circles and “cooler” colors (e.g., blue). As can be seen from Figure 3 and Table 1, by far the most frequently occurring GDP was at SFO for low ceilings. The second most frequently occurring GDP was for winds at EWR, followed by low ceilings at PHL. The condition “Other” appearing in Fig. 3 was explicitly observed in numerous NTML GDP records with no additional information to indicate the actual cause of the GDP.



**Figure 3. Frequency, cause, and location of GDPs issued five or more times in 2011.**

**Table 1. Top-10 GDP occurrences for 2011 by location and cause.**

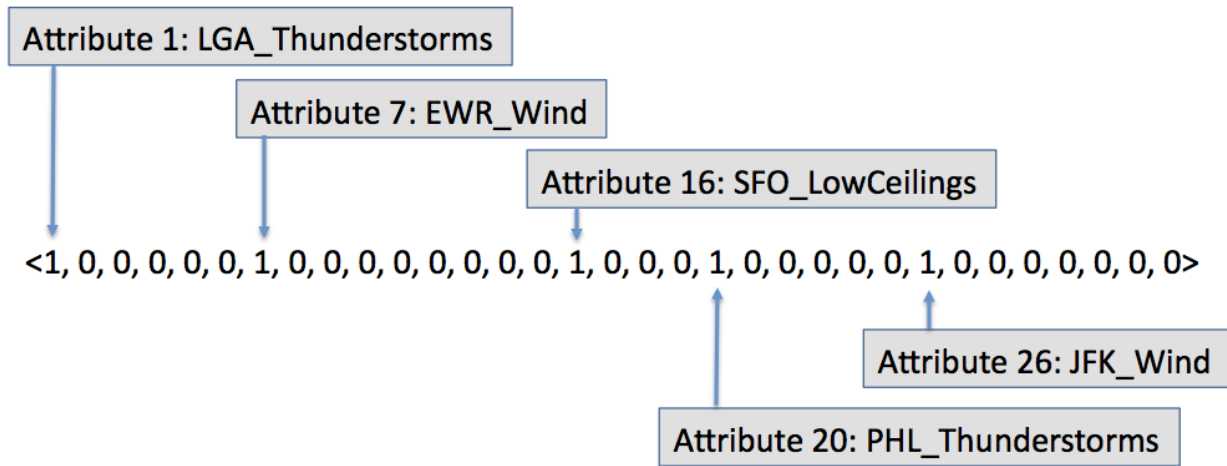
Airport	Condition	Number of GDPs	Airport	Condition	Number of GDPs
SFO	Low Ceilings	166	EWR	Low Ceilings	49
EWR	Winds	71	JFK	Low Ceilings	36
PHL	Low Ceilings	59	LGA	Low Ceilings	34
BOS	Low Ceilings	53	ORD	Low Ceilings	33
JFK	Winds	51	LGA	Thunderstorms	19

For the purpose of identifying similar days in the NAS, from a GDP perspective, each day for which a GDP occurred was characterized in terms of the presence or absence of a GDP at a specific location for a specific cause. To limit the number of features associated with each day, only those GDP locations and causes that occurred at least five times throughout 2011 were considered. With this filtering there were a total of 33 features used to characterize each day, which are listed in order in Table 2. Here the ordering is used when specifying the vector of features associated with each day, as illustrated below.

**Table 2. GDP features used to describe each day.**

Feature Index	Feature Name	Feature Index	Feature Name	Feature Index	Feature Name
1	LGA_Thunderstorms	12	ORD_Snow/Ice	23	DEN_Wind
2	LGA_LowCeiling	13	ORD_Thunderstorms	24	JFK_Thunderstorms
3	LGA_Snow/Ice	14	ORD_LowCeilings	25	JFK_Volume
4	LGA_Wind	15	SFO_LowVisibility	26	JFK_Wind
5	LGA_LowVisibility	16	SFO_LowCeilings	27	JFK_LowCeilings
6	EWR_Snow/Ice	17	SFO_Wind	28	MSP_Snow/Ice
7	EWR_Wind	18	PHL_LowCeilings	29	JAX_Other
8	EWR_LowVisibility	19	PHL_LowVisibility	30	BOS_LowVisibility
9	EWR_Thunderstorms	20	PHL_Thunderstorms	31	BOS_LowCeilings
10	EWR_LowCeiling	21	PHL_Wind	32	ATL_Thunderstorms
11	ORD_Wind	22	DFW_Thunderstorms	33	DAY_Other

Here, “LGA\_Thunderstorms” for example would be used to specify the presence or absence of GDP at LGA due to thunderstorms. A representation of the vector of 33 features used for describing August 19, 2011 is shown in Figure 4.



**Figure 4. Vector of features describing August 19, 2011 with GDPs labeled.**

Using the feature names and indices from Table 2, this vector of features indicates that there were GDPs at LGA and PHL for thunderstorms (features 1 and 20), EWR and JFK for winds (features 7 and 26) and low ceilings at SFO (feature 16) on August 19, 2011.

**B. Airport-level Clusters**

At each of the airports for which the Expectation Maximization clustering algorithm was used to verify the stated cause of the GDP, the following eight features were considered:

- Hourly scheduled arrival rates from the FAA’s Aviation System Performance Metric (ASPM) system
- Hourly METAR observations, which included the wind speed, wind gust, wind direction, ceilings, visibility and a textural description of any precipitation that may be impacting the airport
- A Boolean value indicating the presence or absence of a GDP during the specified hour

Since each of these eight features may or may not be relevant for the clustering at a particular airport for a particular cause, a feature relevance analysis (see Section II.B) was conducted using the Weka software package prior to

clustering the data. It is worth noting that the list of possible features considered in this initial study is by no means exhaustive, and there are numerous additional features, such as winds aloft, downstream volume constraints, etc., that will be examined in future studies. Section IV.B will describe in more detail the features that were relevant for each airport.

## IV. Results

### A. National-level Clustering Results

Figures 5 through 8 depict 4 of the 18 national-level airport GDP clusters for 2011. The color and size of the circles in each image are used to distinguish the cause and frequency of the GDPs. Here larger sized circles indicate more GDPs for a particular cause. Starting with Figure 5, this cluster contains 24 days and is characterized by the presence of low ceiling related GDPs in BOS and throughout the New York area. Cluster 3, which is illustrated in Fig. 6, also contains 24 days and is characterized by GDPs due to low ceilings at SFO and wind-related GDPs at EWR. Cluster 5, is illustrated in Fig. 7, depicts a cluster containing 19 days that is dominated by GDPs due to low ceilings at ORD. Finally, Fig. 8 depicts Cluster 9, which contains 73 days during which low ceilings impacted SFO. It is worth noting that only 318 days were clustered using this approach, since 47 days in 2011 were free of GDPs at the U.S. airports listed in Table 2 for the specified causes.

A textual description of the major causes and locations of the GDPs associated with each of the national-level clusters is presented in Table 3. As seen from this table, the clusters with the most elements were clusters 7 and 9. Cluster 7 contained 57 members that were characterized in terms of thunderstorm-related GDPs at ORD and PHL, snow/ice related GDPs at MSP, low ceilings EWR and LGA, and volume related GDPs at JFK. Cluster 9 is dominated by days for which low ceiling GDPs impacted SFO. The smallest clusters were Clusters 1 and 17 both of which contain a single day. On each of these days, a fairly unique number of GDP locations and causes impacted the NAS.

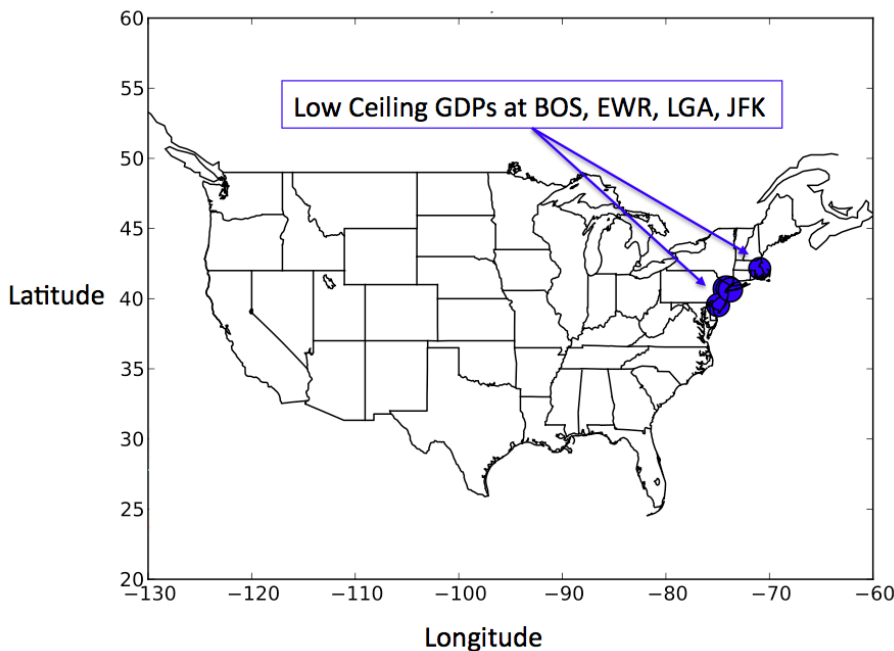


Figure 5. Cluster 2 with low ceiling GDPs at BOS and the New York area and 24 members (days).



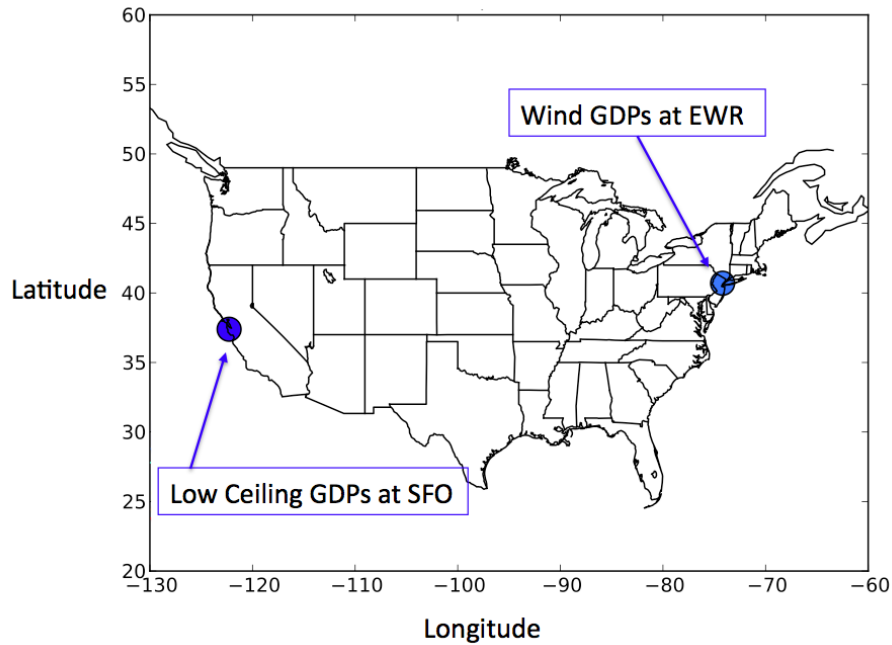


Figure 6. Cluster 3 with low ceiling GDPs at SFO and wind-related GDPs at EWR and 24 members (days).

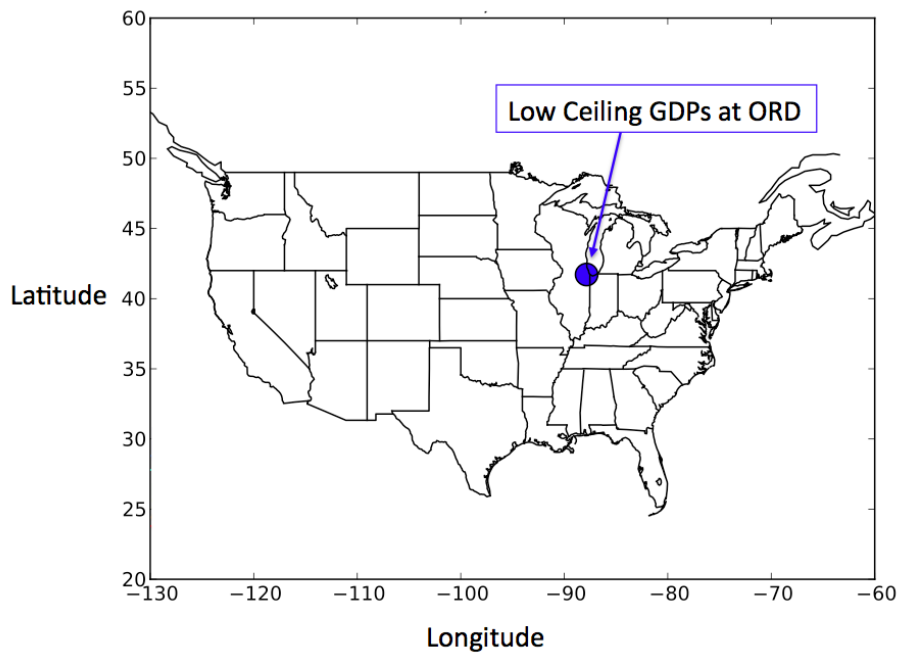
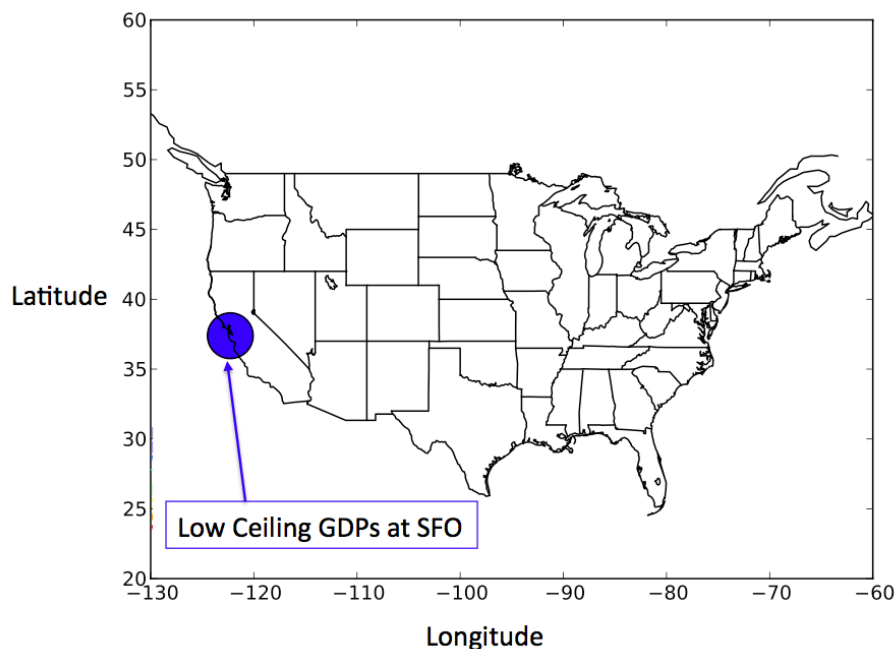


Figure 7. Cluster 5 with low ceiling GDPs at ORD and 19 members (days).



**Figure 8. Cluster 9 with low ceiling GDPs at SFO and 73 members (days).**

In general, the clustering approach proposed in this study creates distinctive and meaningful national-level clusters using historical GDP data. In future work, this clustering approach will be extended to support “day-of-operation” planning, where a given day will be classified in terms of one of the available clusters. One of the major challenges associated with this extension will be using weather forecasts, such as the TAF, to determine if a GDP is required at a specific airport. Elements that will make this challenging include: (1) uncertainty in the need for a GDP due to errors in the weather forecast and (2) correctly identifying where the weather constraint exists (e.g., at the runway, on final approach, or in the en route area). The clustering work presented in the next subsection is an initial step in this direction, although these results still rely on the use of current observations, not forecasts.

**Table 3. Major causes and locations of GDPs associated with each national-level cluster for 2011.**

Cluster Index	Number of Members	Description	Cluster Index	Number of Members	Description
0	7	Low ceilings at SFO and PHL with winds at EWR, LGA, and JFK	9	73	Low ceilings at SFO
1	1	GDPs of various causes throughout the U.S.	10	10	Thunderstorms at LGA and EWR
2	24	Low ceiling at BOS, EWR, LGA, and JFK	11	21	Winds at ORD, EWR, LGA, and JFK
3	24	Low Ceilings at SFO, winds at EWR and thunderstorms at ATL and ORD	12	6	Thunderstorms at DFW
4	3	Low visibility at PHL	13	17	Winds at EWR and LGA with low ceilings at SFO
5	19	Low ceilings at ORD	14	30	Winds at EWR and PHL
6	12	Low ceilings at PHL and BOS, low visibility at EWR and LGA and thunderstorms at JFK	15	2	Low ceiling in the New York area and ORD
7	57	Thunderstorms at ORD and PHL, snow/ice at MSP, low ceilings at EWR and LGA and volume at JFK	16	2	Low ceilings in the New York area, BOS and SFO
8	9	Thunderstorms at ATL	17	1	Thunderstorms at DFW, LGA, and JFK with low ceilings at SFO, EWR, and BOS

## B. Airport-level Clustering Results

Using available METAR, ASPM and GDP start and end time records from NTML, the Expectation Maximization algorithm in the Weka software package was used to cluster these hourly records for the airports most frequently impacted by GDPs in 2011 in an attempt to verify the NTML stated cause of the GDPs. This section presents sample results for the following airports: SFO, EWR and ATL. Here SFO and EWR were selected because there were more GDPs implemented at these airports than any other airports in the U.S. in 2011. Additionally, the weather constraints associated with the SFO and EWR GDPs were fairly localized at the airports (e.g, low ceilings for SFO and winds at EWR). In contrast, the ATL GDPs were associated with “thunderstorms,” so this was expected to be a more challenging scenario for the clustering algorithm because of the en route impacts of the thunderstorms. Future studies will consider additional airports that are commonly impacted by GDPs in the U.S.

### 1. Low Ceiling GDPs at San Francisco International Airport

There were a total of 166 low ceiling related GDP days at SFO in 2011. Applying the feature relevance analysis, which was described in Section II.B, to the eight available features indicates that the scheduled number of arrivals, the ceilings and the visibility were the most relevant factors for identifying low ceiling GDP events at SFO. With these features the five clusters shown in Table 4 were identified. A total of 11% of the 8,760 hourly records were in Cluster 0, 23% were in Cluster 1, 5% were in Cluster 2, 10% were in Cluster 3 and 50% were in Cluster 4. Note that these some of these hourly records were associated with a GDP, while others were not. The distribution of hourly records with and without a GDP is shown in the last two rows of Table 4. In principle, given a weather and traffic forecast the clusters presented in Tables 4 through 6 can be used to determine if a GDP is going to be required for a particular hour on a given day. Naturally weather and demand forecast uncertainties will complicate this procedure, but future extensions of this work are underway to use data mining techniques to support day-of-operations GDP planning.

The EM clustering algorithm determined that the most likely cluster associated with a low ceiling GDP was Cluster 2, while the cluster least likely to be associated with a low ceiling GDP was Cluster 4. Cluster 2 is characterized by a relatively high scheduled arrival rate (20.9 aircraft/hr) with mean ceilings of 1,508 ft. For reference, marginal visual flight rules occur when the ceilings range from 1,000 to 3,000 ft. Similarly, the cluster least likely to be associated with a low ceiling GDP event at SFO is Cluster 4, which is characterized by a relatively high arrival rate (29.6 aircraft/hr), mean ceilings of 3,193 ft and visibility of 10 mi. So, in general the clustering results appear to agree with the NTML stated cause for low ceilings events at SFO. It is also worth noting that Clusters 1 and 3 are associated with early morning and late night operations at SFO, as indicated by the low mean scheduled arrival rates associated with each of these clusters.

A closer look at the distribution of the hourly GDP and non-GDP events (bottom two rows in Table 4) highlights one of the major challenges associated with applying data mining techniques to historical GDP records. The problem that is being dealt with is very much akin to attempting to find the proverbial “needle in the haystack.” In this case, a significant number of the hourly GDP records do in fact reside in Cluster 2, however over 80% of the hourly GDP records actually reside in Cluster 4. The EM algorithm associates Cluster 4 with non-GDP events, since over 3,600 hours of non-GDP events reside within this cluster. However, it is clear from the results that under current day operations there are situations in which similar traffic and weather scenarios lead to GDPs while at other times no GDP is implemented.

**Table 4. SFO Clusters.**

Mean Feature Value	Cluster 0	Cluster 1	Cluster 2 (GDP Likely)	Cluster 3	Cluster 4 (GDP Unlikely)
Scheduled Arrivals (ac/hr)	29.3	2.8	20.9	3.6	29.6
Ceilings (ft)	18,330	987	1,508	10,666	3,193
Visibility (mi)	9.7	10.0	6.5	10.0	10.0
Hourly GDP Events	3	41	113	0	800
Hourly Non-GDP Events	955	1,997	351	884	3,616

## 2. Wind GDPs at Newark International Airport

A total of 71 days in 2011 contained wind-related GDP events at EWR. The feature relevance analysis indicated that the wind speed, wind gust, wind direction, ceiling and the number of scheduled arrivals were the most relevant factors for identifying the presence of these GDPs. With these five factors, the four clusters shown in Table 5 were produced. A total of 28% of the 8,760 hourly GDP records were in Cluster 0, 20% of the records were in Cluster 1, 36% were in Cluster 2 and 17% were in Cluster 4. Based on the clustering, the algorithm determined that the most likely cluster associated with a wind-related GDP at EWR was Cluster 3, which is characterized by a relatively high number of scheduled arrivals (28.7 aircraft/hr on average), high sustained winds and high wind gusts. Similarly, the cluster most associated with the absence of a wind-related GDP at EWR was Cluster 2. For this cluster, the scheduled arrivals again are high (30.0 aircraft/hr on average), there are no gusting winds and the wind speed is relatively low. As in the SFO low ceilings scenario, these clustering results again make intuitive sense, and indicate that the stated cause of the wind-related GDPs at EWR are well supported by the weather observations at the airport.

**Table 5. EWR Clusters.**

Mean Feature Value	Cluster 0	Cluster 1	Cluster 2 (GDP Unlikely)	Cluster 3 (GDP Likely)
Wind Speed (kts)	4.2	6.5	7.3	13.8
Wind Gust (kts)	0.1	0.2	0	21.6
Wind Direction (deg.)	21.0	247.8	220.8	248.9
Scheduled Arrivals (ac/hr)	17.4	4.6	30.0	28.7
Ceiling (ft)	7,226	9,689	8,135	6,530
Hourly GDP Events	35	1	268	318
Hourly Non-GDP Events	2,379	1,717	2,904	1,138

### 3. Thunderstorm GDPs at Atlanta International Airport

There were thunderstorm related GDPs at ATL on 18 days in 2011. A feature relevance analysis indicated that the wind speed, wind gust, precipitation type and the number of scheduled arrivals were the most relevant factors for predicting these GDPs. Intuitively, additional factors, such as cross winds, are important when identifying thunderstorm related GDPs. The fact that wind direction was not highly correlated with thunderstorm related GDPs at ATL is likely due to the low number of thunderstorm GDPs at ATL, and additional en route weather effects that need to be accounted for when predicting the presence of a thunderstorm related GDP.

With these four factors, the four clusters shown in Table 6 were identified. A total of 54% of the 8,760 hourly GDP records were in Cluster 0, 4% of the records were in Cluster 1, 8% were in Cluster 2 and 33% were in Cluster 4. Based on the clustering, the algorithm determined that the most likely cluster associated with a thunderstorm-related GDP at ATL was Cluster 1, which is characterized by a relatively high number of scheduled arrivals (65.8 aircraft/hr on average), low winds and either rain or drizzle at the airport. Similarly, the cluster most associated with the absence of a thunderstorm-related GDP at ATL was Cluster 0. For this cluster, the scheduled arrivals again are high (77.5 aircraft/hr on average) and there are low winds and no precipitation impacting the airport. Although thunderstorm-related GDPs at ATL are relatively scarce, as compared to low ceiling GDPs at SFO, the METAR observations at the airport do appear to confirm the stated cause of these GDPs.

**Table 6. ATL Clusters.**

Mean Feature Value	Cluster 0 (GDP Unlikely)	Cluster 1 (GDP Likely)	Cluster 2	Cluster 3
Wind Speed (kts)	7.1	5.9	12.5	5.3
Wind Gust (kts)	2.9	0.2	15.23	0
Scheduled Arrivals (ac/hr)	77.5	65.8	54.4	7.1
Precipitation	None	Rain/Drizzle	None/Rain	None
Hourly GDP Events	43	15	10	1
Hourly Non-GDP Events	4,698	361	719	2,913

Based on the three examples presented in this subsection, it does appear as if the stated causes of historical GDPs can be substantiated with historical METAR records. In future studies, this work should be extended to consider both forecasts at the airport (e.g., TAF reports) and weather in the vicinity of the airport, since the weather hazards leading to the GDP often occur in the proximity of the impacted airport instead of directly at the airport.

## V. Conclusions

“Typical” days in the National Airspace System (NAS) are identified in this study based on the locations and causes of Ground Delay Programs (GDPs). A k-means clustering algorithm is used to identify 18 unique clusters for this purpose. Each day in 2011 was characterized in terms of 33 features, which were a combination of the GDP causes and locations that were historically implemented. The features were Boolean variables, either yes or no depending on whether or not a GDP occurred at a particular airport for a particular cause. Because of this, the Jaccard index was adopted as the distance metric for cluster members, as opposed to a more conventional distance measure like the Euclidean distance. In general the clusters generated by the algorithm appeared distinct and meaningful. By far the largest cluster consisted of 73 days on which GDPs at San Francisco International Airport (SFO) were the primary GDPs observed in the NAS.

In an effort to verify that the stated causes of the GDPs could be substantiated with available weather observations, an Expectation Maximization clustering algorithm was applied to the 8,760 hourly Meteorological Aerodrome Report, scheduled arrival rate, and GDP start and end time records for SFO, ATL and EWR. For all three airports, clusters where GDPs were likely and unlikely to occur were identified. For example, a cluster with low ceilings, low visibility and high scheduled arrival rates was identified as a cluster likely associated with GDPs at SFO, which is to be expected given the seasonal impacts of the marine stratus layer at SFO. However, similar weather and traffic conditions often lead to situations where a GDP was implemented under some situations but not others. Additionally, the list of features associated with each airport is by no means exhaustive, and additional features such as the season, the winds aloft, en route convection, etc. will be considered in follow-on studies.

Findings from this initial study indicate that it is possible to identify similar days under which the National Airspace System operated, and clustering techniques appear to be promising methods for identifying the major causes of Ground Delay Programs.

### Acknowledgments

The authors will like to thank Mr. Rich Jehlen from the Federal Aviation Administration for motivating and supporting this work.

### References

- <sup>1</sup>Federal Aviation Administration, *Facility Operation and Administration*, Chap. 17-5, U.S. Department of Transportation, 2010.
- <sup>2</sup>Krozel, J., "Survey of Weather Impact Models used in Air Traffic Management," *AIAA Aviation Technology, Integration, and Operations (ATIO) Conference*, AIAA 2010-9020, Sept. 2010.
- <sup>3</sup>Rios, J., Jehlen, R., and Zhu, Z., "A Spatial Database for Reroute Planning," *IEEE/AIAA 31<sup>st</sup> Digital Avionics System Conference (DASC)*, Oct. 2012.
- <sup>4</sup>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 (accepted).
- <sup>5</sup>Ward, J. H., "Hierarchical Grouping to Optimize and Objective Function," *Journal of American Statistical Association*, Vol. 58, No. 301, pp. 236-244, 1963.
- <sup>6</sup>Callahan, 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," *4<sup>th</sup> USA/Europe Air Traffic Management R&D Seminar*, 2001.
- <sup>7</sup>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.
- <sup>8</sup>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.
- <sup>9</sup>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.
- <sup>10</sup>Han, J. and Kamber, M., *Data Mining Concepts and Techniques*, Morgan Kaufman Publishers, San Francisco, CA, 2001.
- <sup>11</sup>Cunningham, J., Cook, L., and Provan, C., "The Utilization of Current Forecast Products in a Probabilistic Airport Capacity Model," *92<sup>nd</sup> Annual American Meteorological Society Meeting*, New Orleans, LA, Jan. 2012.
- <sup>12</sup>Wang, Y., "Prediction of Weather Impacted Airport Capacity using RUC-2 Forecasts," *IEEE/AIAA 31<sup>st</sup> Digital Avionics System Conference (DASC)*, Oct 2012.
- <sup>13</sup>Do, C. B. and Batzoglou, S., "What is the expectation maximization algorithm?," *Nature Biotechnology*, Vol. 26, No. 8, Aug. 2008.
- <sup>14</sup>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.