Assessing Terminal Weather Forecast Similarity for Strategic Air Traffic Management

Yi Liu*, Michael Seelhorst, Alexey Pozdnukhov, Mark Hansen
Institution of Transportation Studies
University of California, Berkeley, CA 94720
liuyisha@berkeley.edu

Michael O. Ball
Robert H. Smith School of Business
University of Maryland, College Park, MD 20742

***Abstract*—In this paper, we propose a semi-supervised learning algorithm to assess weather forecast similarity for strategic air traffic management. Instead of tweaking the distance manually, the distance metric between weather forecasts is supervised by and automatically learnt from the similarity and dissimilarity relationships pre-defined by the users. The distance metric considers the difference in each weather variable and also the interaction between two weather variables' differences. Using the proposed methodology, two case studies are performed at Newark Liberty International Airport (EWR), where historically similar days in 2011 are identified for two given days-of-operation. The results show that similar weather forecasts could lead to very different airport acceptance rate and runway configuration outcomes in terms of capacity profiles and selections of runway configuration. Since differences in different weather phenomena are weighted differently in the distance metric, the algorithm could produce similar days to a given day which have considerable differences in some unimportant weather phenomena. If we use past decisions on similar days to inform a GDP decision for the given days-of-operation, assuming similar demand profiles, the analysis would have provided beneficial suggestions.**

Keywords-Air Traffic Management, Decision Making, Similar Days, Data Mining Technique, Terminal Weather Forecast

I. INTRODUCTION

Air traffic managers today are typically limited to personal experience to make Traffic Flow Management (TFM) decisions [1, 2]. These decisions include whether or not there is a need for Traffic Management Initiatives (TMIs), such as Ground Delay Program (GDP) and Airspace Flow Program (AFP), and how TMIs should be planned when they are considered necessary. Managers with different experiences or different preferences may create different TMI plans for the same situation. This unpredictability in decision creates uncertainty for National Airspace System (NAS) users and may hinder them from taking effective proactive actions. To address this issue, systematic approaches should be developed to better inform and assist managers in TFM decision making.

One way of achieving this goal is to provide capacity profiles based on similar historical days to TFM decision makers. As illustrated in Fig. 1, two pieces of information are needed for traffic managers to make a TFM decision: a demand profile, which is known on a given day-of-operation; and a set of capacity profiles, which are to some degree uncertain. There is considerable research literature concerning the use of capacity profiles in the TMI planning process [3-7] and, to a lesser extent, the generation of these capacity profiles.

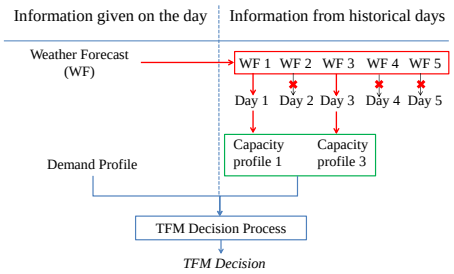

Figure 1. Flowchart of Traffic Flow Management Decision Making

Reference [8] classifies historical capacity profiles into a small number of nominal scenarios for a given airport by using *K*-means clustering, which does not incorporate weather forecast. Reference [9] generates different capacity profiles for San Francisco International (SFO) airport for a given day by using empirical distributions of weather forecast error. Reference [10] creates capacity profiles also for SFO by using cumulative distribution functions of fog clearance time estimated from historical data. In [9] and [10], the source for weather information is exclusive to SFO and the models simplify the capacity profile by assuming an Airport Acceptance Rate (AAR) of 30 arrivals per hour before the fog clearance and 60 arrivals per hour afterwards. Reference [11] overcomes these limitations by basing the analysis on publically accessible Terminal Aerodrome Forecasts (TAFs) and developing possible capacity profiles from historical capacity scenarios. In this work, they find similar historical days to a given day using *K*-means clustering or Dynamic Time Warping (DTW) based on TAF.

In our work, we borrow the logic of generating capacity profiles in [11], which is shown in the upper part of Fig 1. First, similar historical days are identified to a given day by assessing the similarity between the weather forecasts. Then the AAR time series of the historically similar days are used as the candidate capacity profiles.

As mentioned, in [11], two ways to identify similar days to a given day are proposed: TAF clustering and DTW. In the former, the day is classified into one of several TAF clusters which has the shortest Euclidean distance between its centroid and the given day's TAF. The days in that cluster are then the similar days. In the DTW approach, the similarity of a given

day TAF to all the historical TAFs is measured and the probability of a capacity profiles is inversely proportional to the degree of similarity. In both methods, the weather phenomena, such as visibility and ceiling, are weighted the same in the distance metrics. Moreover, while there are many plausible ways of defining the distance metrics (Euclidean, city block, etc.), the choice of distance metric employed in this work is rather arbitrary.

This research addresses these issues by applying a semi-supervised learning algorithm to measure similarity between weather forecasts, where the distance metric is automatically learnt from similarity/dissimilarity relationships pre-defined by the users. The metric captures the differences in each weather phenomenon and also the interactions between different weather phenomena. The weights of the squared and the interaction terms are estimated based on the pre-defined similarity/dissimilarity relationships.

The methodology will be described in Section II. In Section III, we describe the data that is used in this analysis. In Section IV, we present results from our case studies at Newark Liberty International Airport (EWR) airport. Finally, we conclude the paper in Section V.

II. METHODOLOGY

In this analysis, we apply a semi-supervised algorithm to identify similar days to a given day by learning distance metric [12]. The proposed algorithm consists of two steps: First, we learn the distance metric between **hourly** weather forecasts. Second, we identify similar days by selecting the days with small total distances in the weather forecast, summing over the hourly distances. In Sections A and B, we will elaborate on the two steps, respectively.

A. Learning Distance Metrics

In this section, we will explain how we generate the distance metric between hourly weather forecasts. Consider learning a distance metric of the form:

$$d_A(WF_i, WF_j) = \|WF_i - WF_j\|_A = \sqrt{(WF_i - WF_j)^T \cdot A \cdot (WF_i - WF_j)} \quad (1)$$

where, $WF_i \in \mathbb{R}^n$ is the weather forecast vector for hour i; n is the dimension of the vector. For this analysis, we use the following pieces of weather information found in the forecasts: ceiling, visibility, with speed and direction, as well as the presence of thunderstorm and snow. In the general case, A is a full matrix with diagonal and off-diagonal distance coefficients. The diagonal coefficients are the weights of the squares of the differences in each weather variable. The off-diagonal coefficients are the weights of the interaction term between two weather variable differences. If we set $A = I$, then the expression becomes the Euclidean distance. If we restrict A to be diagonal, this then corresponds to learning a metric in which the different weather attributes are given different weights without any interaction between them. In all cases, the required constraint on A is that it must be positive semi-definite, $A \succcurlyeq 0$. This ensures d_A to be a metric—satisfying non-negativity and the triangle inequality.

The key step in this approach is finding the A matrix and associated distance metric. We learn a distance metric that respects predetermined similarity between hours. Suppose we know that certain pairs of the hourly WF_i's are similar:

$$S: (WF_i, WF_j) \in S, \text{ if } \mathrm{WF_i} \text{ and } \mathrm{WF_j} \text{ are similar} \quad (2)$$

We can then learn a distance metric d_A respecting this so that similar hours end up close to each other. This is achieved by solving the following optimization problem:

$$\min_A \sum_{(WF_i, WF_j) \in S} \|WF_i - WF_j\|_A^2 \quad (3)$$

$$\text{s.t.} \sum_{(WF_i, WF_j) \in D} \|WF_i - WF_j\|_A \geq 1, \quad (4)$$

$$A \succcurlyeq 0. \quad (5)$$

where, the objective is to minimize the squared distance between the pairs of points in S. Since this is trivially solved with $A = 0$, we add the constraint (5) to ensure that A does not collapse the dataset into a single point. Here, D can be a set of pairs of hourly forecasts known to be dissimilar if such information is available; otherwise, it can be the complement of S. The objective function is linear in the parameters A, and both of the constraints can be verified to be convex. Thus, the optimization problem is convex, which enables us to derive efficient, local-minima-free algorithms to solve it. We consider this algorithm as semi-supervised since it learns a data transformation guided by user-defined S and D matrices as opposed to a fully supervised predictive model targeted at forecasting targets from sample input-output pairs.

The question that remains unanswered is how we define similarity and dissimilarity sets. In other words, we need to find the pairs of hours that belong to the sets S and D. The ultimate goal of identifying weather forecast similarity is to assist in air traffic management decision-making. Towards this goal, we define two hours as similar if all of the following conditions hold:

- The runway configuration is the same;
- Both hours have the same Meteorological Conditions (MC), either Instrument MC (IMC) or Visual MC (VMC);
- The absolute difference in actual AARs is smaller than S_{AAR}, where S_{AAR} is the AAR similarity threshold.

On the contrary, we define two hours to be dissimilar if they have different runway configurations, MCs and the differences in AARs are larger than the dissimilarity threshold, Di_{AAR}.

There is certainly more than one way to generate S and D. More research is needed to explore good definitions of the two sets. The definition of these sets is a place where feedback from TFM decision makers could be incorporated to tailor the similarity assessment to their need. In this paper, we focus on assessing terminal weather forecast similarity for airport capacity profile generation. But the methodology we use has generality and can be applied in other contexts. For instance, similarity in en route weather could be assessed with S and D defined using en route network characteristics such as traffic density.

B. Identifying Similar Days

In this section, using the learnt distance metric between hourly weather forecasts from Section A, we will describe how

similar days are identified for a given day. Assume the concerned time horizon is from hour T_s to hour T_e. One way of defining the distance of weather forecast between two days is to sum up the squares of the hourly distances with the time horizon:

$$D_{J,K} = \sum_{i=T_s}^{T_e} \left\| WF_{J,i} - WF_{K,i} \right\|_A^2 \quad (6)$$

where, $D_{J,K}$ is the total distance between day *j* and day *k*; $WF_{J,i}$ is the weather forecast in hour *i* on day *J*; $WF_{K,i}$ is the weather forecast in hour *i* on day *k*. For a given day, similar days are then days with small total distances. The user can decide on the number of similar days, *Ns*, to look at.

Once historically similar days are identified following the proposed mechanism, the time series of the actual runway configuration, MCs, and AARs on these days may then be provided to traffic managers as references for decision making.

III. DATA

The analysis uses data from two sources: Aviation System Performance Metrics (ASPM) and Terminal Aerodrome Forecast (TAF). ASPM report provides hourly data on runway configuration, MC and AAR.

Historical TAFs provides hourly data for terminal weather forecasts. The authors developed a Matlab script to download the historical TAF information from www.ogimet.com. Each TAF was read in as a text file. Afterwards, a parser written in Matlab was used to convert the text files to user-friendly numerical data. Weather variables created based on the TAF data are listed below.

First, we have two indicator variables representing the presence of thunderstorms—*TS*,and snow—*Sn*. Visibility in the forecasts ranges from 0.25 miles to 7 miles. We expect the effect of the visibility to be much stronger as it approaches zero, also when the value is below 4 as opposed to above 4, where 4 is the threshold for visual approach conditions. We use two variables for visibility, including a natural log transform—*Vis1* = log(visibility) and a discontinuity to capture the nonlinear effects—*Vis2* = max(0,log(visibility/4)). Ceiling in the forecasts ranges from 100 feet to 25,000 feet. Similar to visibility, two variables are defined for ceiling: *Ceil1* = log(ceiling) and *Ceil2* = max (0,log(ceiling/3000)), where 3000 feet is the threshold of visual approach condition. The TAFs contain both wind speed and direction. We include three variables to capture this information. When wind direction is specified (some observations have variable wind direction), we decompose the wind speed into two components: speed from the North to the South and speed from the East to the West. The direction of the wind blowing from north to south and from east to west is considered positive. When the wind direction is variable, we consider wind by its absolute speed only. This gives us three wind variables:

- *Ws*: absolute wind speed. It equals to the reported wind speed when wind direction is unspecified, and zero when wind direction is specified
- *WN*: component of wind speed from the North to the South. It equals to zero if wind direction is unspecified
- *WE*: component of wind speed from the East to the West. It equals to zero if wind direction is unspecified

The weather forecast vector, *WF*, is then a 9-dimensional vector:

$$WF = [TS; Sn; Vis1; Vis2; Ceil1; Ceil2; Ws; WN; WE] \quad (7)$$

For each day, a new TAF is updated every two to three hours. Therefore, we should determine which forecast to use before assessing weather forecast similarity between a given day and the historical days. Since the purpose of identifying similar days is to assist in traffic management decision-making, one way to select the TAF is to refer to the traffic management decision time. At the decision time, we compare the most recently issued TAF on the given day to the most recent TAFs that were available at this time for the historical days.

Each TAF forecasts weather for 24 hours from the forecast issuance time. The number of hours that will be compared to identify similar days is determined by the planning horizon.

For the case study, we pulled ASPM and TAF hourly data for EWR for years 2011 and 2012. TAF data is missing for 2011 November. We thus removed the observations for November from the ASPM data as well.

IV. EWR CASE STUDIES

In this section, we apply the proposed methodology to terminal weather forecast similarity assessment at EWR. This airport is selected since it is one of the most congested airports in the US and has many GDPs. Two case studies are performed for two selected days-of-operation: September 20, 2012 and June 8, 2012. There was no GDP on September 20, 2012 and there was a GDP on June 8, 2012 due to wind. The GDP was planned to start at 4:47 pm and end at 9 pm but actually ended at 5:43 pm.

We first use the ASPM and TAF data from 2011 to generate the *A* matrix. Then, using the learnt distance metric, we identify five similar days for each of the two days. When defining the *S* and *D* sets, we set the thresholds for similarity and dissimilarity in AARs as 1 and 7 arrivals per hour respectively. The results for the A matrices and similar days are shown and discussed in the following two subsections.

A. Similar Days for Sep 20, 2012

There was no GDP on 9/20/2012. Many of the EWR GDPs start around noon (without specification, time is assumed as local time) and planned for around 10 hours on average [13]. Accordingly, we set the time horizon of the analysis from noon to midnight. All the analysis results are then based on hourly data between noon and midnight from ASPM and TAF, including estimation of the *A* matrix. Moreover, the most recent TAF that was available at noon is referenced for values of the weather variables.

The estimation results for the *A* matrix are shown in Table I, containing values of diagonal and off-diagonal elements. The diagonal entries are distance coefficients (also referred to as weights) for the squares of the differences in each weather variable. The off-diagonal entries are distance coefficients for the interaction terms of two weather variables' differences.

To explain the interpretation of the *A* matrix, we offer a simple example. For a pair of hours, assume all the other weather conditions are the same but there could be differences in thunderstorm and snow conditions. Then the distance metric is reduced to $61.15\cdot(\Delta TS)^2 + 7.05\cdot(\Delta Sn)^2 + 2\times4.78\cdot\Delta TS\cdot\Delta Sn$, where ΔTS and ΔSn are the thunderstorm and snow differences between these two hours. Since thunderstorm happens more often in the summer season and it snows only in the winter, the interaction term $\Delta TS\cdot\Delta Sn$ almost always takes the value -1 or 0. Therefore, the distance metric can take three values:

- 61.15 if one hour has thunderstorm, the other hour does not have thunderstorm, and they have the same snow condition (presumably no snow);
- 7.05 if one hour has snow, the other hour does not have snow, and they have the same thunderstorm condition (presumably no thunderstorm);
- $61.15+7.05-2\times4.78$ if they have different thunderstorm and snow conditions.

These indicate that there is a large difference between a thunderstorm hour and a non-thunderstorm hour, whereas there is less difference between a snow hour and a non-snow hour. If one hour has thunderstorm and the other hour has snow, the distance between these two hours is smaller than two hours with different thunderstorm conditions but the same snow conditions. This indicates how the interaction term affects the distance between 2 hours in which snow and thunderstorm conditions are both different.

TABLE I. ESTIMATION RESULTS ON A MATRIX

Variables	***TS***	***Sn***	***Vis2***	***Vis1***	***Ceil2***	***Ceil1***	***Ws***	***WN***	***WE***
TS	61.15[a]	4.78	4.14	15.52	-11.74	12.30	2.22	-0.47	-0.56
Sn	4.78	7.05	20.28	-7.67	-3.80	3.84	-0.10	-0.07	-0.10
Vis2	4.14	20.28	66.55	-17.37	-29.90	30.09	2.91	-0.75	-0.81
Vis1	15.52	-7.67	-17.37	39.05	-40.69	41.07	8.46	-1.35	-1.31
Ceil2	-11.74	-3.80	-29.90	-40.69	87.12	-87.76	-15.24	2.64	2.61
Ceil1	12.30	3.84	30.09	41.07	-87.76	88.41	15.35	-2.66	-2.63
Ws	2.22	-0.10	2.91	8.46	-15.24	15.35	2.76	-0.47	-0.46
WN	-0.47	-0.07	-0.75	-1.35	2.64	-2.66	-0.47	0.08	0.08
WE	-0.56	-0.10	-0.81	-1.31	2.61	-2.63	-0.46	0.08	0.08

a. All the weights are scaled by a factor of 1000.

The diagonal coefficients are the weights of the squared terms of each weather variable's differences. But they cannot be interpreted as the relative importance of the squared differences to the distance since the values of the weather variable differences vary. In order to compare the contributions of each weather phenomenon to the distance, we create five hypothetical *A* matrices for the five types of weather phenomena based on the full *A* matrix: thunderstorm, snow, visibility, ceiling and wind. The hypothetical *A* matrix for a weather phenomenon is created by keeping the diagonal distance coefficient(s) of the weather variables belonging to this weather phenomenon and, where applicable, their interaction coefficients, and replacing the rest of the entries with zero. Following this, there will be 1 non-zero element for thunderstorm and snow respectively, 4 non-zero elements for visibility and ceiling respectively, and 9 non-zero elements for wind, as highlighted in boxes in Table I. Hypothetical distances between the hourly TAFs are then estimated for each pair of hours selected for this analysis using the five hypothetical *A* matrices. For each weather phenomenon, these hypothetical distances could also be estimated by using the full *A* matrix and assuming no differences in the variables representing the other weather phenomena. Now, we can compare the contributions of different phenomena to the distance by comparing the values of these hypothetical hourly distances.

In Table II, we present summary statistics for the hypothetical hourly distances for the five weather phenomena. The median value is the highest for wind, followed by ceiling, whereas the rest of the median values are zero. These results indicate that typically wind differences account for most of the distance and ceiling differences are responsible for the rest. This is mainly because wind differences and ceiling differences are common whereas more than half of the pairs of hours have the same values for thunderstorm, snow and visibility variables, according to the percentages of non-zeroes. If we only consider non-zero observations for each weather phenomena, the average distances due to thunderstorm differences are the largest on average. This indicates that two hours would be viewed as the most different when one has thunderstorm and the other one does not. The average distances due to visibility and ceiling differences are similar in magnitude, and much higher than the distance due to wind differences. Out of the five differences, wind differences have the smallest average impact when considering only the non-zero values. As a result, even though wind difference is the most common difference, the average distance induced by this difference across both zero and non-zero observations is smaller than the average distance generated by the visibility differences and ceiling differences (as shown in Row 1 in Table II), which contribute significantly to the distance metric when they occur, and they do a fair amount of the time.

TABLE II. STATISTICS OF HYPOTHETICAL HOURLY DISTANCES

Statistics	**Thunderstorm**	**Snow**	**Visibility**	**Ceiling**	**Wind**
Mean	3.5[a]	2	47	94.6	24.5
Median	0	0	0	3.4	20.1
Max	247.3	84	491.5	809.9	320.8
% non-zero obs.	1.43	2.41	31.3	71.6	99.6
Mean of non-zero obs.	247.3	84	150.3	132.1	24.6
Median of non-zero obs.	247.3	84	149	5.5	20.2
Std. of non-zero obs.	0	0	103.4	190.9	21.8

a. All the weights are scaled by a factor of 1000.

In order to visualize the similarity between each hour, we have applied a metric Multi-Dimensional Scaling (MDS) [14] to the pair-wise distance matrix. An MDS plot provides a distance-preserving visualization of the data such that the pairwise distances in 9-dimensional space are reproduced in 2 dimensions with minimum distortion. It helps analyzing the effect of the metric transform learned by the algorithm and the scaling applied to original variables. An example of the MDS plot is presented in Fig. 2 for *Ceil1*—log(ceiling), where the color scale corresponds to the values of the original *Ceil1* variable. As shown, the ceiling value roughly increases from the right to the left. Hours with similar ceiling values are clustered to some degree. We have considered similar plots for all the variables in MDS projection. They are not shown here but the information is conveyed in Fig. 2. The plot of visibility has similar trend as ceiling, where the value increase from the right to the left. Hours with thunderstorms and snow locate at

the bottom and on the top, respectively, which are far away from the majority of the points. This indicates that a pair of hour with different thunderstorm or snow conditions is more different from that with different ceiling or visibility conditions. In the plots of wind variables, there is no obvious trend in the spatial distribution of hours with different wind speeds and wind components. This indicates that the contribution of wind difference to the hourly distance is similar in the range of wind speed we have here. The cluster of good weather conditions is enlarged in the plot. They are hours with high ceiling, high visibility, no thunderstorm, and no snow.

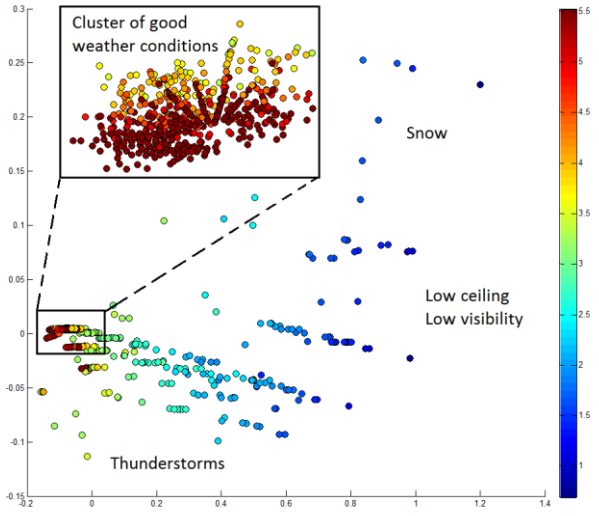

Figure 2. Metric Multi-Dimensional Scaling Results

Using (3), the similar days from 2011 were identified for Sep 20, 2012, based on the estimated *A* matrix and values of the TAF variables. The five similar days are selected as the five days with the smallest total hourly distances in TAF compared to the given day. All the hours were VMC except for the hour from 5 pm to 6pm on Sep 20, 2012. The actual AARs, selected runway configurations and the GDP decisions of the similar days and the given day are summarized in Table III. The first 4 similar days all share the same runway configuration with the given day for a considerable duration, whereas 12/3/2011 has a totally different configuration. The AARs of the first two similar days are similar to the given day, especially in terms of sum of AARs. The AARs of the last three similar days are smaller than those of the given day for about half of the time, with the largest hourly difference as 10 arrivals per hour.

TABLE III. ACTUAL OBSERVATIONS ON 09/20/2012 AND ITS HISTORICALLY SIMILAR DAYS

	Given Day 9/20/2012 No GDP		Similar Day 1 9/16/2011 No GDP		Similar Day 2 7/27/2011 No GDP	
Hour (GMT[b])	**AAR**	**Rwy Conf.**	AAR	Rwy Conf.	AAR	Rwy Conf.
9/20/16Z	**46**	**4R, 11|4L**	46	4R, 11|4L	48	4R, 11|4L
9/20/17Z	**46**	**4R, 11|4L**	46	4R, 11|4L	48	4R, 11|4L
9/20/18Z	**46**	**4R, 11|4L**	46	4R, 11|4L	48	4R, 11|4L
9/20/19Z	**46**	**4R, 11|4L**	46	4R, 11|4L	48	4R, 11|4L
9/20/20Z	**41**	**4R | 4L**	46	4R, 11|4L	38	4R | 4L
9/20/21Z	**38**[a]	**4R | 4L**	46	4R, 11|4L	38	4R | 4L
9/20/22Z	**38**	**4R | 4L**	46	4R, 11|4L	38	4R | 4L
9/20/23Z	**46**	**4R, 11|4L**	46	4R, 11|4L	38	4R | 4L
9/21/00Z	**44**	**4R | 4L**	46	4R, 11|4L	38	4R | 4L
9/21/01Z	**38**	**4R | 4L**	46	4R, 11|4L	38	4R | 4L
9/21/02Z	**38**	**4R | 4L**	46	4R, 11|4L	38	4R | 4L
9/21/03Z	**38**	**4R | 4L**	46	4R, 11|4L	38	4R | 4L
	Similar Day 3 10/30/2011 GDP		Similar Day 4 7/9/2011 No GDP		Similar Day 5 12/3/2011 No GDP	
Hour (GMT)	AAR	Rwy Conf.	AAR	Rwy Conf.	AAR	Rwy Conf.
9/20/16Z	36	4R | 4L	42	4R | 4L	38	22L | 22R
9/20/17Z	36	4R | 4L	48	4R, 11 | 4L	38	22L | 22R
9/20/18Z	38	4R | 4L	41	4R | 4L	38	22L | 22R
9/20/19Z	38	4R | 4L	38	4R | 4L	38	22L | 22R
9/20/20Z	38	4R | 4L	38	4R | 4L	38	22L | 22R
9/20/21Z	38	4R | 4L	38	4R | 4L	38	22L | 22R
9/20/22Z	38	4R | 4L	38	4R | 4L	38	22L | 22R
9/20/23Z	38	4R | 4L	38	4R | 4L	38	22L | 22R
9/21/00Z	38	4R | 4L	38	4R | 4L	38	22L | 22R
9/21/01Z	38	4R | 4L	38	4R | 4L	38	22L | 22R
9/21/02Z	38	4R | 4L	38	4R | 4L	38	22L | 22R
9/21/03Z	38	4R | 4L	38	4R | 4L	38	22L | 22R

a. This hour was IMC where the rest were all VMC.

b. GMT time is 4 hours ahead of local time during daylight saving and 5 hours ahead otherwise.

The forecasted ceiling, wind speed (*Ws*) and Wind direction (*Wdir*) for the given day and similar days are summarized in Table IV. There were no thunderstorms or snow and visibility was 7 mile for all the hours. The weather conditions are very similar for these days except for wind. The wind directions in the similar days are very different from those on the given day, except 12/3/2011. The overall wind speed is generally small for all days. This indicates that the difference in wind direction is not very important in determining similarity when wind speed is not high. The weather conditions on 12/3/2011 are similar to the given day. However, the AAR and runway configuration of 12/3/2011 and 6/8/2012 are different according to Table III. This indicates that capacity profiles and selection of runway configuration could be very different given similar terminal weather forecasts. This could be a result of inherent uncertainty in weather forecasts, or non-weather factors (such as demand and facility outages) that also influence the runway configuration and AAR.

On similar day 3, a GDP was planned from 3 pm to 10 pm local time. There were no GDPs for the other 4 similar days. Therefore, if the demand profile for the given day is similar to that from the historically similar days, the similarity analysis would suggest no GDP on the given day, which was the actual TFM decision.

TABLE IV. TAF FOR 09/20/2012 AND ITS HISTORICALLY SIMILAR DAYS

	Given Day 9/20/2012			Similar Day 1 9/16/2011			Similar Day 2 7/27/2011		
Hour (GMT)	**Ceiling**	***Ws***	***Wdir***	Ceiling	*Ws*	*Wdir*	Ceiling	*Ws*	*Wdir*
9/20/16Z	**40**	**9**	**100**	250	12	340	250	11	340
9/20/17Z	**40**	**9**	**100**	250	12	340	250	11	340
9/20/18Z	**40**	**9**	**100**	250	9	320	250	11	340
9/20/19Z	**250**	**8**	**140**	250	9	320	250	10	310
9/20/20Z	**250**	**8**	**140**	250	9	320	250	10	310
9/20/21Z	**250**	**8**	**140**	250	9	320	250	10	310
9/20/22Z	**250**	**8**	**140**	250	9	320	250	8	300
9/20/23Z	**250**	**8**	**140**	250	9	320	250	8	300
9/21/00Z	**250**	**5**	**110**	200	7	330	250	8	300
9/21/01Z	**250**	**5**	**110**	200	7	330	250	4	300
9/21/02Z	**250**	**5**	**110**	200	7	330	250	4	300
9/21/03Z	**250**	**5**	**110**	200	7	330	250	4	300

	Similar Day 3 10/30/2011			Similar Day 4 7/9/2011			Similar Day 5 12/3/2011		
Hour (GMT)	Ceiling	*Ws*	*Wdir*	Ceiling	*Ws*	*Wdir*	Ceiling	*Ws*	*Wdir*
9/20/16Z	250	14	330	250	13	330	250	7	30
9/20/17Z	250	14	330	250	13	330	250	7	30
9/20/18Z	250	14	330	250	13	320	250	6	140
9/20/19Z	250	14	330	250	13	320	250	6	140
9/20/20Z	250	13	320	250	13	320	250	6	140
9/20/21Z	250	13	320	250	13	320	250	6	140
9/20/22Z	250	13	320	250	13	320	250	6	140
9/20/23Z	250	13	320	250	13	320	250	6	140
9/21/00Z	250	13	320	250	13	320	250	6	140
9/21/01Z	250	6	320	250	8	340	250	6	140
9/21/02Z	250	6	320	250	8	340	250	6	140
9/21/03Z	250	6	320	250	8	340	250	6	140

B. Similar Days for June 8, 2012

There was a GDP planned on 6/8/2012 from 4:47 pm to 9 pm due to wind. The program was actually ended earlier at 5:43 pm. To find similar days for assisting in decision-making, we set the time horizon as 4 pm to 10 pm. All the analysis results are then based on hourly data on this time horizon from ASPM and TAF, including estimation of the *A* matrix. The most recent TAF that was available at this time is referenced for values of the weather variables. It is worth mentioning that the *A* matrix is different from before because we are using a different set of pairs of hours and different TAFs. Usually, the most recent TAFs prior to noon and 4 pm are issued around 10 am and 1:30 pm, respectively. The results for this case study are shown in Tables V to VIII. There were no thunderstorms or snow and visibility was 7 miles for all the days.

As shown in Tables V and VI, compared to the previous case, snow and ceiling differences are weighted more where the other three differences are weighted less.

TABLE V. ESTIMATION RESULTS ON A MATRIX

Variables	***TS***	***Sn***	***Vis2***	***Vis1***	***Ceil2***	***Ceil1***	***Ws***	***WN***	***WE***
TS	19.46	32.41	8.61	16.16	-53.73	53.02	5.42	-2.12	-1.46
Sn	32.41	54.03	13.18	27.53	-88.80	87.64	8.99	-3.51	-2.41
Vis2	8.61	13.18	27.74	-5.05	-34.19	33.42	2.92	-1.24	-0.84
Vis1	16.16	27.53	-5.05	19.81	-38.65	38.30	4.21	-1.59	-1.10
Ceil2	-53.73	-88.80	-34.19	-38.65	155.17	-153.03	-15.27	6.04	4.14
Ceil1	53.02	87.64	33.42	38.30	-153.03	150.93	15.07	-5.96	-4.08
Ws	5.42	8.99	2.92	4.21	-15.27	15.07	1.52	-0.60	-0.41
WN	-2.12	-3.51	-1.24	-1.59	6.04	-5.96	-0.60	0.24	0.16
WE	-1.46	-2.41	-0.84	-1.10	4.14	-4.08	-0.41	0.16	0.11

a. All the weights are scaled by a factor of 1000.

TABLE VI. STATISTICS OF HYPOTHETICAL HOURLY DISTANCES

Statistics	**Thunderstorm**	**Snow**	**Visibility**	**Ceiling**	**Wind**
Mean	5.9	7.7	43.6	128.7	34.5
Median	0	0	0	7.7	28.6
Max	139.5	232.4	458	894.5	278.9
% non-zero obs.	4.19	3.33	35.8	74.4	99.5
Mean of non-zero obs.	139.5	232.4	121.7	173	34.6
Median of non-zero obs.	139.5	232.4	108.3	10.7	28.3
Std. of non-zero obs.	0	0	85.3	247.2	27.3

a. All the weights are scaled by a factor of 1000.

The capacity profiles and the runway configurations are similar between the given day and the first four similar days, as shown in Table VII. But again, differences are observed in the weather forecasts. The AARs and runway configurations are not very similar for the given day and similar day 5, neither are the weather forecasts. Specifically, an additional arrival runway (Runway 11) was used for much of the time on similar day 5, which greatly increased the arrival capacity. Further study is required to determine if this might have been predicted based on the weather forecast.

Out of the five similar days, only 8/1/2011 had a GDP. If the demand profiles are similar between 6/8/2012 and the similar days, then the analysis suggests no GDP for this day. Although there was a GDP for the given day, it was cancelled less than 1 hour after its implementation. The suggestion would have been helpful since it appears a GDP was not necessary on the given day.

TABLE VII. ACTUAL OBSERVATIONS ON 06/08/2012 AND ITS HISTORICALLY SIMILAR DAYS

	Given Day 6/8/2012 GDP		Similar Day 1 8/1/2011 GDP		Similar Day 2 12/30/2011 No GDP	
Hour (GMT)	**AAR**	**Rwy Conf.**	AAR	Rwy Conf.	AAR	Rwy Conf.
6/8/21Z	**39**	**22L | 22R**	32	22L | 22R	38	22L | 22R
6/8/22Z	**38**	**22L | 22R**	32	22L | 22R	38	22L | 22R
6/8/23Z	**38**	**22L | 22R**	35	22L | 22R	38	22L | 22R
6/9/00Z	**38**	**22L | 22R**	35	22L | 22R	38	22L | 22R
6/9/01Z	**38**	**22L | 22R**	35	22L | 22R	42	22L | 22R
6/8/21Z	**38**	**22L | 22R**	35	22L | 22R	42	22L | 22R

	Similar Day 3 7/23/2011 No GDP		Similar Day 4 12/31/2011 No GDP		Similar Day 5 5/1/2011 No GDP	
Hour (GMT)	AAR	Rwy Conf.	AAR	Rwy Conf.	AAR	Rwy Conf.
6/8/21Z	38	22L | 22R	38	22L | 22R	52	11, 22L | 22R
6/8/22Z	38	22L | 22R	46	11, 22L | 22R	52	11, 22L | 22R
6/8/23Z	38	22L | 22R	38	11, 22L | 22R	52	11, 22L | 22R
6/9/00Z	38	22L | 22R	38	22L | 22R	49	22L | 22R
6/9/01Z	38	22L | 22R	38	22L | 22R	38	22L | 22R
6/8/21Z	38	22L | 22R	38	22L | 22R	38	22L | 22R

TABLE VIII. TAF FOR 06/08/2012 AND ITS HISTORICALLY SIMILAR DAYS

	Given Day 6/8/2012			Similar Day 1 8/1/2011			Similar Day 2 12/30/2011		
Hour (GMT)	**Ceiling**	***Ws***	***Wdir***	Ceiling	*Ws*	*Wdir*	Ceiling	*Ws*	*Wdir*
6/8/21Z	**250**	**12**	**270**	250	10	260	250	6	190
6/8/22Z	**250**	**12**	**270**	250	10	260	250	6	190
6/8/23Z	**250**	**12**	**270**	250	10	260	250	6	190
6/9/00Z	**250**	**12**	**270**	250	10	260	250	6	190
6/9/01Z	**150**	**12**	**300**	250	8	290	250	5	160
	Similar Day 3 7/23/2011			Similar Day 4 12/31/2011			Similar Day 5 5/1/2011		
Hour (GMT)	Ceiling	*Ws*	*Wdir*	Ceiling	*Ws*	*Wdir*	Ceiling	*Ws*	*Wdir*
6/8/21Z	150	12	280	40	9	290	200	10	160
6/8/22Z	150	12	280	40	9	290	200	10	160
6/8/23Z	150	12	280	40	9	290	200	10	160
6/9/00Z	150	12	280	40	9	290	200	10	160
6/9/01Z	250	10	280	250	10	290	150	7	160

V. CONCLUSIONS AND FUTURE WORK

In this work, we propose a semi-supervised algorithm for assessing weather forecast similarity for air traffic management. The distance metric between hourly TAFs is automatically learnt from similarity and dissimilarity relationships pre-defined by comparing the actual outcomes (AAR, runway configuration and MC) for the hours. Distance coefficients for the squared differences in the weather variables and the coefficients for the interaction between two weather variables' differences are estimated. Then distance between two days is calculated as the sum of the squared hourly distances over a given time horizon. Finally, the degree of similarity of a historical day to a given day is inversely proportional to the distance between them.

Using the proposed algorithm, we perform two case studies at EWR, where historically similar days from 2011 are identified for 9/20/2012 and 6/8/2012 respectively. The distance metric shows that wind difference is the most common weather difference at EWR. However, on average, ceiling and visibility differences contribute most to the hourly TAF distances. A day with severe thunderstorms or heavy snow has the largest TAF distance compared to a good weather day.

Comparing the AAR and runway Configuration outcomes, and TAFs of the given day to the similar days, we learn that similar weather forecasts can lead to different outcomes, which demonstrate that the uncertainty in TFM might be unavoidable given the inherent uncertainty in weather forecast. It is also observed that wind direction could be very different for similar days when wind speed is small. This indicates that difference in the wind direction may not be important in determining the outcomes, when the wind speed is below a certain threshold. More work should be done to find this threshold.

Summarizing out two case studies, there was no GDP on 9/20/2012, and for its five historically similar days, only one day had GDP. There was a GDP planned for five hours on 6/8/2012 but was cancelled four hours earlier. Out of the historically similar days for 6/8/2012, only one day had GDP. Assuming similar demand profiles, then the analysis would suggest no GDP in either case based on the past decisions, and this would indeed have probably been the better course of action given the early cancellation on 6/8/2012.

This work is still at its preliminary research stage. In the on-going work, we are improving the current method and exploring other approaches to similarity assessment under the same idea of learning distance metric. One alternative way to supervise is defining similarity/dissimilarity relationship based on the realized performance such as delay cost, rather than actual AAR, runway configuration and MC. In this case, the distance metric will be learned for a pair of days instead of a pair of hours. Moreover, this will change the flowchart shown in Fig. 1 because demand profile will also be used in calculating delay cost.

REFERENCES

[1] S. R. Wolfe, J. L. Rios, "A method for using historical ground delay programs to inform day-of-operation programs," AIAA Gidance, Navigation, and Control Conference, 2011.

[2] S. Grabbe, B. Sridhar, A. Mukherjee, "Similar days in the NAS: an airport perspective," Aviation Technology, Integration, and Operations Conference, 2013.

[3] O. Richetta, A. R. Odoni, "Dynamic solution to the ground holding problem in air traffic control," Transportation Research Part A, Vol. 28, pp. 167-185, 1994.

[4] A. Mukherjee, M. Hansen, "A dynamic stochastic model for the single airport ground holding problem," Transportation Science, Vol. 41, pp. 444-456, 2007.

[5] B. M. Ball, R. Hoffman, A. Mukherjee, "Ground delay Program planning under uncertainty based on the ration-by-distance principle," Transportaion Science, Vol. 44, pp. 1-14, 2010.

[6] H. D. Sherali, J. M. Hill, M. V. McCrea, A. A. Trani, "Integrating slot exchange, safety, capacity, and equity mechanisms within an airspace flow program," Transportation Science, Vol. 45, pp. 271-284, 2011.

[7] Y. Liu, M. Hansen, "Incorportating predictability into cost optimization for ground delay program," unpublished.

[8] P. B. Liu, M. Hansen, A. Mukherjee, "Scenario-based air traffic management: from theory to practice," Transportation Research Part B, Vol. 42, pp. 685-702, 2008

[9] L. S. Cook, B. Wood, "A model for determining ground delay program parameters using a probabilistic forecast of stratus clearing," Proceedings of 8th USA/Europe Air Traffic Management R&D Seminar, 2009.

[10] A. Mukjerjee, M. Hansen, S. Grabbe, "Ground delay program planning under uncertainty in airport capacity," Transportation Planning and Technology, Vol 35., pp. 611-628, 2012.

[11] G. Buxi, M. Hansen, "Generating day-of-operation probabilistic capacity scenarios from weather forecasts," Transportation Research Part C, Vol. 33, pp. 153-166, 2013

[12] P.X. Eric, A.Y. Ng, M.I. Jordan, S. Russel, "Distance metric learning, with appliation to clustering with side-informtion," Advances in Neural Information Processing Systems, Vol. 15, , pp. 505-512, 2002.

[13] Y. Liu, M. Hansen, "Evaluation of performance of ground delay program," Transportation Research Record, in press.

[14] J. B., Kruskal, M. Wish, "Multidensional scaling," Sage Univerity Paper Series on Quantitative Application in the Social Science, pp. 7-11, 1978