Generating Day-of-Operation Probabilistic Capacity Profiles from Weather Forecasts

Mark Hansen, Gurkaran Singh Buxi
Department of Civil & Environmental Engineering
University of California, Berkeley
Berkeley, California, USA
mhansen@ce.berkeley.edu, gkb@berkeley.edu

***Abstract*—It is common understanding that weather plays an important role in determining the capacity of an airport. Severe weather causes capacity reductions, creating a capacity demand imbalance, leading to delays. The role of air traffic flow management (ATFM) measures is to reduce these delay costs by aligning the demand with the capacity. Ground delay program (GDP) is one such measure. Though the GDP is initiated in poor weather conditions, and weather forecasts are subject to errors, present GDP planning procedures are essentially deterministic in nature. Forecast weather is translated into deterministic capacity predictions on which GDP planning is based. Models which employ probabilistic capacity profiles for planning GDPs have been developed, but their application has been limited by the inability to create such profiles from weather forecasts. This paper focuses on San Francisco International Airport (SFO) and provides a methodology to generate probabilistic capacity profiles from the two terminal weather forecasts: Terminal Aerodrome Forecast and San Francisco Marine Initiative (STRATUS). The profiles are inputs to a static stochastic GDP model to simulate ATFM strategies. The solution from the model is evaluated against realized capacities to determine the benefit of the forecast. The benefit of inclusion of the weather forecast is assessed by comparing costs of delays from ATFM strategies simulated from probabilistic profiles developed without the weather forecasts. It is also shown that inclusion of weather forecasts reduces the cost of delays. The paper also compares the cost of delays from strategies simulated using the profiles generated from TAF and STRATUS. It is shown that on average TAF offers similar benefit in controlling cost of delay when compared to STRATUS, indicating that other airports would also benefit from using TAF in planning of operations.**

Keywords- Air traffic flow management; Ground delay program; Probabilsitc Capacity Profiles; K-means Clustering; Terminal Aerodrome Forecast; STRATUS

I. INTRODUCTION

Adverse weather conditions in the vicinity of an airport reduce the operational capacity of the airport leading to an imbalance between capacity and demand. This capacity-demand imbalance creates en-route airspace congestion leading to delays, an increase in cost and elevating Air Traffic Flow Management (ATFM) risk. If adverse weather is present in the vicinity of the airport, the Federal Aviation Administration (FAA) implements Ground Delay Programs (GDPs). GDPS mitigate weather-induced airspace congestion by metering the arrival of aircraft to the destination airport. The metering matches the number of flights arriving in a period with the "airport acceptance rate" (AAR) forecast. The metering of flights is achieved by delaying inbound flights on ground at the origin airport prior to their departure. If the AAR forecast is perfectly accurate, the metering from the GDP ensures that the total delay costs are minimized.

It is common understanding that the AAR is primarily influenced by the weather in the vicinity of the airport and thus AAR forecasting necessitates a terminal weather forecast. The weather forecasts are seldom accurate in perfectly predicting the conditions and can thus lead to inaccurate predictions of the AAR. GDP models found in the literature incorporate the uncertainty in the AAR and can be classified in two broad categories: dynamic models and static models. In dynamic models as the information is updated, ground holding decisions are revised, incorporating a wait-and-see strategy. Dynamic models require scenario trees to represent the uncertainty in the AAR. Conversely, in a static model, decisions made once are not revised. Static models require probabilistic capacity profiles as inputs. Reference [1] contains more details on the types of GDP models. A substantial amount of theoretical research on GDP models has been found. However, capacity profiles or scenario trees that the models assume are often used for illustrative purposes and are hence not considered to be a part of ATFM decision making. There exists a gap in literature on the development of specific day-of-operation probabilistic capacity profiles, hindering incorporation of tailor-made day-of-operation strategies in ATFM. This paper focuses on the development of probabilistic capacity profiles generated from day-of-operation weather forecast. The potential benefit of incorporating weather forecasts into day-of-operation would lead to a more accurate decision making, lower costs and reducing ATFM risk.

This paper primarily uses San Francisco International Airport (SFO) as a case study due to its uniqueness in employing two sources of terminal weather forecasts: Terminal Aerodrome Forecast (TAF) and the SFO Marine Stratus Forecast System (STRATUS). TAF is a weather forecast issued for all major airports and is updated four times in a day. It contains forecast information on visibility, ceiling etc for the entire day. Unlike other major airports, SFO experiences a low altitude marine stratus cloud layer during the summer which reduces the airport capacity. STRATUS forecasts the "burn-off" time of these marine clouds i.e. the time when the capacity would increase. We construct probabilistic capacity profiles from both types of terminal forecasts.

The contribution of this paper is that it provides a methodology which uses several statistical techniques to convert weather forecasts into specific day-of-operation probabilistic capacity profiles. These profiles are provided as inputs in a Static Stochastic GDP model to simulate ATFM strategies. We have also developed a test-bed, assessing the benefit under the different methods of probabilistic profile generation against perfect information. The test-bed also provides an opportunity to score the benefits of the profiles generated by the two weather forecasts. The benefit on inclusion of the forecast is gauged by comparing the cost of the strategies developed with and without the weather forecast. Lastly, including the day-of-operation weather forecast would assist the air traffic controllers and dispatchers in planning a schedule for arrivals at the beginning of the day. The schedule could also be modified if new forecasts are available with time.

We have focused on the days when predictions from STRATUS were available i.e. marine clouds were observed in the terminal area. We construct probabilistic profiles from 7am to 10pm as the bulk of the traffic is observed in this duration. Since this paper deals with SFO, it may limit our generality of finding but this by no means is a limitation as the methodology can be applied to any other major airport for which the TAF is issued.

This paper proceeds as follows. Section II provides the relevant literature review. Section III describes the weather forecasts and the generation of probabilistic profiles by various methods. Section IV presents the GDP model and a comparative comparison of the costs of the strategies obtained from the profiles developed in Section III. Section V offers conclusion.

II. LITERATURE REVIEW

The current National Airspace System (NAS) rarely incorporates uncertainty of the weather forecasts into tactical decisions. Operations planning assume a deterministic approach using expected weather conditions [2]. Since it is difficult to accurately predict AAR, several researchers have formulated GDP models which require either stochastic capacity profiles or scenario trees for AAR as inputs [1,3,4]. This paper focuses on generation of stochastic profiles and details on scenario trees can be found in [5]. The generation of stochastic scenario found in the literature is primarily applied to stochastic programs in finance [6]. In the aviation community, a stochastic scenario for the capacity an airport refers to a set of several stochastic capacity profiles of the AAR. A stochastic profile is a time series of AAR values with a particular probability of occurrence. For a given airport there are several profiles depicting different possible evolution of capacity. Thus stochastic profiles capture the uncertainty in the future capacities.

Reference [5] formulates a methodology for developing stochastic profiles from historical AAR data for various airports in the United States. The profiles are the centroids of the clusters obtained after K-means clustering the AAR time series. Their approach in profile construction is devoid of any weather forecast information. Reference [7] has focused on the STRATUS forecast thus its application seems to be limited to San Francisco International Airport (SFO). They assume a sharp capacity increase when the stratus "burns-off" which is not representative of the actual increase. They use the day of operation STRATUS forecast to determine the optimal end time of the GDP by modeling the time of the capacity increase as a random variable. Reference [8] gauged the imprecision of the forecast weather information with the actual weather by comparing delays. They compare the delays under the forecast conditions and actual weather conditions by associating a fractional loss of capacity under different weather conditions. Their approach ignores the imprecision in the forecast, which when implemented in ATFM simulation leads to a higher cost.

There is a gap in the research to develop specific day-of-operation capacity profiles or profiles based on weather forecasts.

III. WEATHER DATA AND PROFILE GENERATION

This section provides methodology for developing the set of profiles $\{S_p\}$ and their probabilities $\{P_p\}$ using various statistical techniques for both types of weather forecasts. With each technique a discussion is also provided to determine the number of profiles. An explanation of the relevant weather forecast precedes the development of profiles. Since we have focused on San Francisco International Airport, we also present a brief case study for the airport.

A. San Francisco International Airport

San Francisco International Airport is the second busiest airport in the California and the tenth busiest airport in the United States. It is also a major hub for United Airlines and a transfer terminal for international travel. The current runway configuration at SFO consists of four runways. The runways are positioned in parallel in two pairs. Two of them are in the north-south direction and the other two are the east-west direction. The separation between the parallel runways is 750ft. In good weather, simultaneous landings can take place on parallel runways accommodating a maximum arrival rate of 60 planes an hour. In bad weather, the landings are restricted to one runway which can reduce the capacity to 30 planes per hour. The 01R-19L is the smallest runway (2280m) while the 28R-10L is the airports' longest runway (3610m). When the 28 runways are in configuration, 28R is normally used for landings.

The San Francisco Bay and the Pacific Ocean are in close proximity to SFO and during the summer months marine stratus clouds (fog) are prevalent over the terminal area. During these months, the fog sets in by early morning and burns off by late morning or early afternoon. The marine stratus clouds play a critical role in determining the airport capacity. When the stratus clouds are below 3500ft in the vicinity of SFO, the airport operates in Instrumental Meteorological Conditions (IMC). The fog prohibits the use of the dual landing and the capacity of SFO is reduced leading to a lower AAR. If the fog coincides with the morning arrival push, it creates a demand capacity imbalance, formulation of en-route airspace queues and delays. On such days the Ground Delay Program is initiated at SFO and its initial planned duration is the burn off time (or improved weather conditions) plus two hours. Due to the strategic importance of SFO and the effect of weather on operations there are two different weather forecast systems available. These predict different yet relevant, to aviation, meteorological conditions. Due to this unique nature of SFO we have used it as a case study. Both the forecasts are used to construct day-of-operation probabilistic profiles.

In the next subsections we describe the various methodologies for profile generation.

B. Naïve Clustering

This method of profile generation does not incorporate any weather forecast and is similar to [5]. In [5], the authors implement K-means clustering to generate the profiles by averaging the AAR time series in the same cluster. Let $[\mathrm{A}]_{\mathrm{T}\times\mathrm{N}} = [\mathrm{A}_1, \mathrm{A}_2, .., \mathrm{A}_\mathrm{N}]$ be the data matrix of the AAR profiles for N days where A_k is column vector of the AAR profile for day k. K-means clustering splits the data matrix into a predefined number of clusters, $\boldsymbol{l}$, where each cluster c_k contains d_k days. The days which have similar AAR profile vectors are grouped together i.e. they are in the same cluster. The similarity is defined as the Euclidean norm between the AAR profile vectors. A smaller value of the Euclidean norm implies greater similarity. After the K-means operation we obtain

$$\{A_h^1\}_{h=1}^{d_1},\ \{A_h^2\}_{h=1}^{d_2},\ \{A_h^3\}_{h=1}^{d_3},\ ,\ \{A_h^l\}_{h=1}^{d_l}$$

such that $\sum_{j=1}^{l} \mathrm{d}_\mathrm{j} = \mathrm{N}$ and

$$\bigcup_{j=1}^{l}\{A_h^j\}_{h=1}^{d_j} = \{A_h\}_{h=1}^{N} \ \textbf{\textit{and}}\ \bigcap_{j=1}^{l}\{A_h^j\}_{h=1}^{d_j} = \boldsymbol{\Phi} \qquad (1)$$

$\{\}$is defined as a set

The optimal number of clusters, $\boldsymbol{l}^*$, is an open problem and there are ad-hoc procedures which assist in determining it. More clusters enable tailor made ATFM strategies for each cluster but each cluster would have a lower probability of occurrence. Reference [5] provides an algorithm involving the pseudo F value to determine the optimal number of clusters.

Procedures like the pseudo F value and Silhouette value measure the compactness of a cluster with respect to other clusters and report an average value over all clusters. The pseudo F value is implemented in SAS and works well with uncorrelated variables (eg: variables obtained after Principle Component Analysis) [9]. The pseudo-F statistic captures the “tightness” of clusters, and is a ratio of the mean sum of squares between clusters to the mean sum of squares within a cluster. Higher pseudo F-values indicate tight clustering and imply that the data is well separated or better clustered. Silhouette value, varying between -1 and 1, measures the similarity between an object and the cluster in which it is classified. The indicator of a strong clustering is the average silhouette value close to 1 [10]. The procedure for silhouette value is implemented in MATLAB. Caution should be exercised in monitoring the number of data points falling within each cluster. If they are too many data points within one cluster one might considering breaking it up on the other hand if there are too few days one would tend to merge two clusters together.

The profiles are the within cluster means of the AAR time series in that cluster. Profile $\boldsymbol{S_i}$ is determined by the average of the AAR profiles in the cluster $\boldsymbol{c_i}$

$$\boldsymbol{S_i} = \left\lceil \sum_{h=1}^{d_i} A_h^i \Big/ d_i \right\rceil \qquad \boldsymbol{i} \in \mathbf{1}..\boldsymbol{l}^* \qquad (2)$$

Where **[ ]** is the nearest integer roundup operator.

The probability of the profile $\boldsymbol{S_i}$ is the proportion of days in $\boldsymbol{c_i}$.
$\boldsymbol{P_i} = \boldsymbol{d_i}/\boldsymbol{N} \qquad \boldsymbol{i} \in \mathbf{1}..\boldsymbol{l}^*$

For our data, **N**=446 days and $\boldsymbol{l}^*$ =3. The number of clusters was determined by the highest Psuedo F value.

We call this procedure ***Naïve Clustering*** as it clusters the AAR without any weather information. Fig.1 shows the profiles for SFO for the summer months.

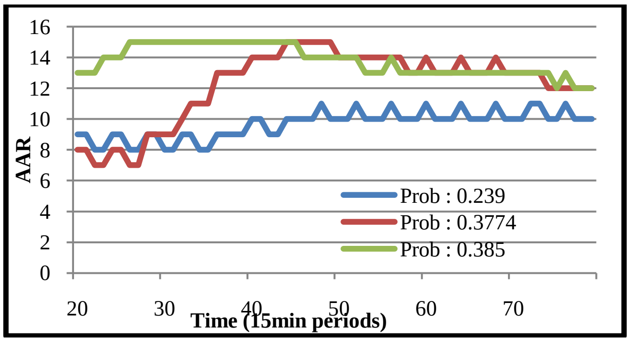

Figure 1. Stochastic Capacity Profiles from Naive Clustering

In the above figure, we observe an oscillating nature of individual profiles. This is because of the way the AAR is reported in the ASPM database. The AAR of 60/hour is split at 15,15,15,15/quarter hour, an AAR of 45/hour is split as 10,11,12,12/quarter hour and an AAR of 30/hour is split as 8,7,8,7/quarter hour leading to an oscillating nature of the profiles when reported in quarter hour intervals.

C. STRATUS and Fog Clustering

STRATUS is a program designed by MIT Lincoln Labs to forecast the time of the fog burn-off. It also includes the probability that the fog would burn off before 17Z, 18Z and 19Z (Z is UTC time) based on the empirical data of all burn-off time forecasts in the same 15 minute time bin. The probabilities are based on an ensemble of statistical models for the weather forecast and atmospheric boundary layer physics models [11]. STRATUS updates the forecast of the burn off time on an hourly basis from 2:00-11:00am PCT. NASA Ames Research Centre maintains a repository where the output from STRATUS is stored for the dates when marine clouds are forecast in the terminal area. For these dates, the data contains the predicted burn off time, its actual burn off time and the probability that the fog would burn-off before 1700, 1800 and 1900 UTC Time.

We consider the STRATUS forecast generated at 8:00am PCT for over 200 days for the summer months (May - September) of 2004-2007. This was chosen as at 8:00am, predictions from the Satellite Statistical Forecast Model (SSFM) in the ensemble become available. We concentrated on the days when the fog burned off between 9:30am and 11:30am PCT as the number of days outside this time bracket were few. These days are binned in 15 minute periods according to the actual fog burn-off time between 9:30 to 11:30 a.m. PCT. In total there are eight fog burn off bins$\{\boldsymbol{B_k}\}_{k=1}^{8}$. The number of days, $\boldsymbol{d_i}$, in bin, $\boldsymbol{B_i}$, is shown in the Table I. From each bin we constructed a probabilistic profile as follows:

$\{\boldsymbol{A_i^{B_k}}\}_{i=1}^{d_k}$ is the set of AAR profiles for the days in bin $\boldsymbol{B_k}$ (k∈1,2,..8).The profile $\boldsymbol{S_i}$ is determined by the average of AAR profiles in $\boldsymbol{B_i}$

$$\boldsymbol{S_i} = \left[\sum_{h=1}^{d_i} A_h^{B_i} \Big/ d_i \right] \quad i \in 1..8$$

(The above equation is similar to (2))

The profiles are shown in the Fig. 2.

TABLE I. BINNING OF ACTUAL FOG BURN-OFF TIME

Bin	Number of Days
9:30-9:45am	15
9:45-10:00am	16
10:00-10:15am	16
10:15-10:30am	11
10:30-10:45am	24
10:45-11:00am	22
11:00-11:15am	18
11:15-11:30am	15

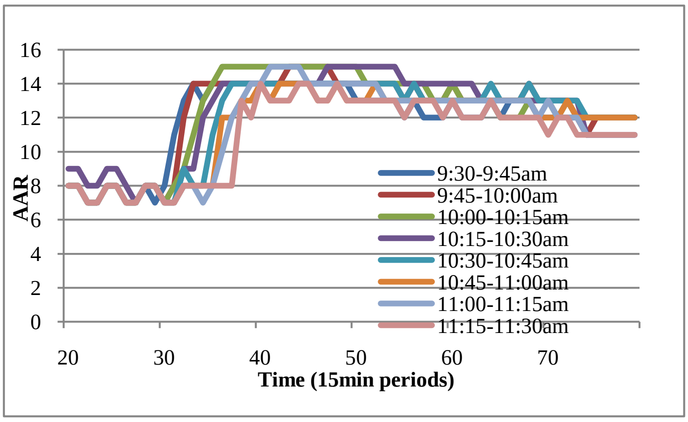

Figure 2. Profiles from Fog Clustering

A closer inspection of the periods when the fog burns off is shown in Fig. 3. It is seen that there is not an immediate increase in the AAR from 30/hour to 60/hour as assumed in [7]. There is a transition period lasting approximately for 45 minutes when the AAR is 45/hour. While calculating the ideal GDP end time, this transition should be taken into account. Ignoring this transition period would lead to an increased cost of airborne delays as the capacity would be over predicted immediately after burn-off. Fig. 3 shows the rise in AAR after burn-off.

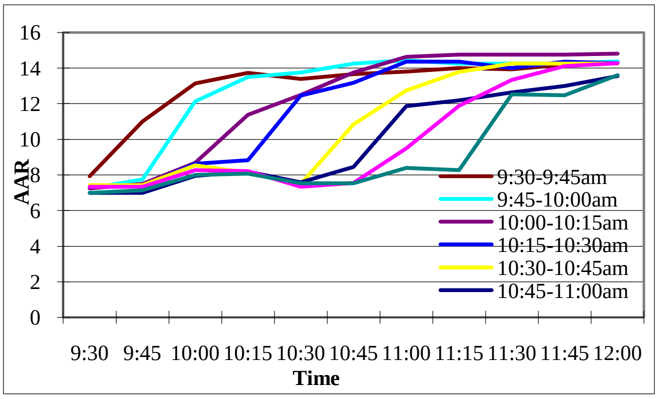

Figure 3. Rise in Capacity at Burn-off

MIT Lincoln labs, on recommendation by the Traffic Management Unit at Oakland center, incorporated "risks" to the output of STRATUS. The "risk" of an hour is the probability that the fog would burn off before that hour. The "risk" is thus a non decreasing function of time. The "risks" can determine the probability of the profiles. Since the risk probabilities are of the form ***P(Burn off < T_1)=P_1***, ***P(Burn off <T_2)=P_2*** and ***P(Burn off <T_3)=P_3*** where $\boldsymbol{T_1 < T_2 < T_3}$, we linearly interpolate the probabilities between the time periods to formulate a Cumulative Distribution Function (CDF) for the fog burn off time.

From the CDF the probability of any bin, $\boldsymbol{B_i}$, can be calculated.

$$\overline{\boldsymbol{P_{B_i}}} = \boldsymbol{Prob}(\boldsymbol{Burn\ off} \leq \lceil \boldsymbol{B_i} \rceil) - \boldsymbol{Prob}(\boldsymbol{Burn\ off} \leq \lfloor \boldsymbol{B_i} \rfloor) \quad \boldsymbol{i} \in [\boldsymbol{1..8}] \qquad (3)$$

Where ⌊·⌋ and ⌈·⌉ are the lower and upper bin boundaries.

Equation (3) establishes the probability of the burn off in a particular bin. Further, if the burn off probability in a particular bin, $\boldsymbol{B_i}$, is $\overline{\boldsymbol{P_{B_i}}}$, then the capacity scenario, $\boldsymbol{S_i}$, depicting burn off in $\boldsymbol{B_i}$, would have a probability

$$\boldsymbol{P_i} = \frac{\overline{\boldsymbol{P_{B_i}}}}{\sum_{i=1}^{8} \overline{\boldsymbol{P_{B_i}}}} \quad \forall \boldsymbol{i} \qquad (4)$$

Equation (4) is a simple renormalization of the probabilities of the bins.

In conclusion, we have generated 8, 15 minute burn-off bins corresponding to the capacity profiles as shown in Fig. 2. From the STRATUS predictions of fog burn- off time for the day-of-operation we can obtain the probabilities for the bins and consequently the probabilities of the profiles.

This methodology translates the STRATUS forecast to build probabilistic capacity profiles. We call this procedure ***Fog clustering***.

D. TAF and Profiles from TAF

The Terminal Aerodrome Forecast (TAF) is a weather forecast issued for every major airport 4 times a day at 6 hours interval by the National Oceanic and Atmospheric Administration (NOAA). It contains the meteorological conditions (wind speed and direction, visibility and cloud type and height) along with qualitative descriptors (rain, fog, mist etc.) for the airport.

A script was written in MATLAB to parse through the data and select TAF data which was issued between 5am and 7am PCT. The TAF issued between this interval assists the SFO control tower, FAA and the airlines to plan the daily operations at SFO during the morning teleconference. In total, we had 446 days when the TAF was issued in this interval. As mentioned earlier the daily operations were considered between 0700-2200hrs PCT (i.e. 60 periods, 1 period = 15 mins). Another script in MATLAB parsed through this filtered data to pick out the relevant weather information: wind, visibility, clouds and their height (wind was broken in two components, wind from north and wind from east direction) for each period. Thus each day was represented by a column vector of length 60 (time periods) ×7 variables (windNorth, windEast, visibility, 4 different clouds height) = 420. Therefore, the entire TAF data set could be represented by a 420 (variables) × 446 (# of days) matrix.

Let $[\mathbf{T}]_{\mathbf{L\times N}}$ be a matrix (N=446, L=420), where T_k is a column representing the TAF for day k. We performed a Principal Component Analysis (PCA) on this matrix. PCA is a standard statistical technique which reduces the dimensionality of the data by converting correlated variables into a smaller number of uncorrelated variables called principal components. The principal components are directions which represent the variation in the data. Thus the first principal component direction represents the maximum variability in the data and each succeeding component accounts for as much of the remaining variability as possible. PCA removes the potential correlation between the forecasted variables for the same day [12]. For example, there might be correlation between visibility and ceiling and also there might be correlation between the forecast weather conditions of adjacent time periods. As a standard preprocessing technique, we normalize the **[T]** matrix i.e. the mean and the variance is 0 and 1 respectively for each variable. Equations (5) through (9) describe the PCA on the data set.

$$[\mathbf{C}]_{\mathbf{L\times L}} = {[\mathbf{T}][\mathbf{T}]^{t}}/{\mathbf{N-1}} \tag{5}$$

Where **[C]** is the empirical correlation matrix.

$$\mathbf{C}\boldsymbol{X} = \boldsymbol{\lambda X} \tag{6}$$

Where lambda is the eigenvalue corresponding to the eigenvector **X**

Sort the eigenvalues in a descending manner (matrix is full rank) i.e.

$$\boldsymbol{\lambda}_{[1]} > \boldsymbol{\lambda}_{[2]} > \boldsymbol{\lambda}_{[3]} > .\,. > \boldsymbol{\lambda}_{[L]} \tag{7}$$

A standard technique is to capture 90% variability, the number of eigenvalues required are

$$\boldsymbol{n} = \boldsymbol{argmax}_{\boldsymbol{k}} \frac{\sum_{p=1}^{k} \boldsymbol{\lambda}_{[p]}}{\sum_{p=1}^{L} \boldsymbol{\lambda}_{[p]}} \leq \mathbf{0.9} \tag{8}$$

Let eigenvector $\mathbf{X}_{[i]}$ correspond to its eigenvalue $\boldsymbol{\lambda}_{[i]}$, then define the matrix $[\mathbf{W}]_{\boldsymbol{n\times L}} = \begin{bmatrix} -\boldsymbol{X}_{[1]}^{T} - \\ \vdots \\ -\boldsymbol{X}_{[n]}^{T} - \end{bmatrix}$

The reduced TAF matrix

$$[\widetilde{\mathbf{T}}]_{\mathbf{n\times N}} = [\mathbf{W}] \times [\mathbf{T}] \tag{9}$$

We proceed to perform a K-means clustering on the matrix$[\widetilde{\mathrm{T}}]$. It has been proved in [13] that performing PCA prior to K-means increases the accuracy of the K-means clustering. The centroids of the clusters are therefore closer to the optimal cluster centroids after PCA.

Thus a K-means clustering on $[\widetilde{\mathrm{T}}]$ with **l** predefined clusters leads to the following

$$\{\widetilde{\boldsymbol{T}}_{\boldsymbol{h}}^{\mathbf{1}}\}_{\boldsymbol{h=1}}^{\boldsymbol{d_1}}, \{\widetilde{\boldsymbol{T}}_{\boldsymbol{h}}^{\mathbf{2}}\}_{\boldsymbol{h=1}}^{\boldsymbol{d_2}}, \{\widetilde{\boldsymbol{T}}_{\boldsymbol{h}}^{\mathbf{3}}\}_{\boldsymbol{h=1}}^{\boldsymbol{d_3}}, \dots, \{\widetilde{\boldsymbol{T}}_{\boldsymbol{h}}^{\boldsymbol{l}}\}_{\boldsymbol{h=1}}^{\boldsymbol{d_l}}$$

such that $$\sum_{j=1}^{l} \boldsymbol{d_j} = \boldsymbol{N} \tag{10}$$

Where d_k is the number of days in the cluster $\boldsymbol{c_k}$

$$\cup_{j=1}^{l}\{\widetilde{\boldsymbol{T}}_{\boldsymbol{h}}^{\boldsymbol{j}}\}_{\boldsymbol{h=1}}^{\boldsymbol{d_j}} = \{\widetilde{\boldsymbol{T}}_{\boldsymbol{h}}\}_{\boldsymbol{h=1}}^{\boldsymbol{N}} \text{ and } \cap_{j=1}^{l}\{\widetilde{\boldsymbol{T}}_{\boldsymbol{h}}^{\boldsymbol{j}}\}_{\boldsymbol{h=1}}^{\boldsymbol{d_j}} = \boldsymbol{\Phi} \tag{11}$$

From this analysis on the TAF issued for SFO, the optimal number of clusters were 2, $\boldsymbol{l}^{*}$=2 i.e. $\boldsymbol{c_1}$ and $\boldsymbol{c_2}$. The number of clusters was determined using the maximum Psuedo F value. Thus, we classify a day in either $\boldsymbol{c_1}$ or $\boldsymbol{c_2}$ depending on the classification of its TAF.

We performed another K-means clustering on the AAR profiles of the days within the c1 and c2. Let $\{\widetilde{\mathrm{A}}_{\mathrm{h}}^{1}\}_{\mathrm{h=1}}^{\mathrm{d_1}}$ be the set of AAR profiles for the days in $\boldsymbol{c_1}$ and $\{\widetilde{\boldsymbol{A}}_{\boldsymbol{h}}^{\mathbf{2}}\}_{\boldsymbol{h=1}}^{\boldsymbol{d_2}}$ be the set of AAR profiles for the days in $\boldsymbol{c_2}$.

The optimal number of AAR clusters within $\boldsymbol{c_1}$ are 2 i.e. $\boldsymbol{c_{1,1}}$ and $\boldsymbol{c_{1,2}}$. Likewise, the optimal number of AAR clusters within $\boldsymbol{c_2}$ are 3 i.e. $\boldsymbol{c_{2,1}}$, $\boldsymbol{c_{2,2}}$ and $\boldsymbol{c_{2,3}}$. Thus $\{\widetilde{\boldsymbol{A}}_{\boldsymbol{h}}^{\mathbf{1,1}}\}_{\boldsymbol{h=1}}^{\boldsymbol{d_{1,1}}}$ and $\{\widetilde{\boldsymbol{A}}_{\boldsymbol{h}}^{\mathbf{1,2}}\}_{\boldsymbol{h=1}}^{\boldsymbol{d_{1,2}}}$ were in cluster $\boldsymbol{c_1}$ and$\{\widetilde{\boldsymbol{A}}_{\boldsymbol{h}}^{\mathbf{2,1}}\}_{\boldsymbol{h=1}}^{\boldsymbol{d_{2,1}}}, \{\widetilde{\boldsymbol{A}}_{\boldsymbol{h}}^{\mathbf{2,2}}\}_{\boldsymbol{h=1}}^{\boldsymbol{d_{2,2}}}, \{\widetilde{\boldsymbol{A}}_{\boldsymbol{h}}^{\mathbf{2,3}}\}_{\boldsymbol{h=1}}^{\boldsymbol{d_{2,3}}}$ in cluster $\boldsymbol{c_2}$.

The highest average Silhouette value determined the number of clusters in this second K-means clustering.

The profiles and their probability are determined as below

$$\boldsymbol{S_{1,1}} = \left[{\sum_{h=1}^{d_{1,1}} \widetilde{\boldsymbol{A}}_{\boldsymbol{h}}^{\mathbf{1,1}}}/{\boldsymbol{d_{1,1}}} \right] \qquad \boldsymbol{P_{1,1}} = {\boldsymbol{d_{1,1}}}/{\boldsymbol{d_{1,1}} + \boldsymbol{d_{1,2}}} \tag{12a}$$

$$S_{1,2} = \left[\sum_{h=1}^{d_{1,2}} \tilde{A}_h^{1,2} \Big/ d_{1,2} \right] \qquad P_{1,2} = d_{1,2} \Big/ d_{1,1} + d_{1,2} \quad (12b)$$

$$S_{2,1} = \left[\sum_{h=1}^{d_{2,1}} \tilde{A}_h^{2,1} \Big/ d_{2,1} \right] \quad P_{2,1} = d_{2,1} \Big/ d_{2,1} + d_{2,2} + d_{2,3} \quad (12c)$$

$$S_{2,2} = \left[\sum_{h=1}^{d_{2,2}} \tilde{A}_h^{2,2} \Big/ d_{2,2} \right] \quad P_{2,2} = d_{2,2} \Big/ d_{2,1} + d_{2,2} + d_{2,3} \quad (12d)$$

$$S_{2,3} = \left[\sum_{h=1}^{d_{2,3}} \tilde{A}_h^{2,3} \Big/ d_{2,3} \right] \quad P_{2,3} = d_{2,3} \Big/ d_{2,1} + d_{2,2} + d_{2,3} \quad (12e)$$

Fig. 4 and Fig. 5 show the profiles and their probabilities for c_1 and c_2.

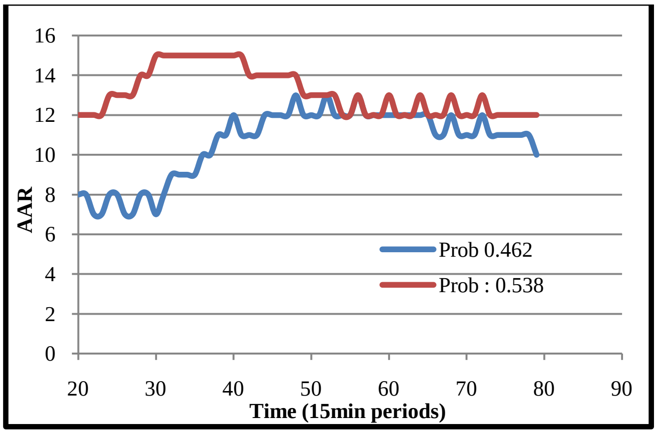

Figure 4. Probabilistic Profiles for cluster 1

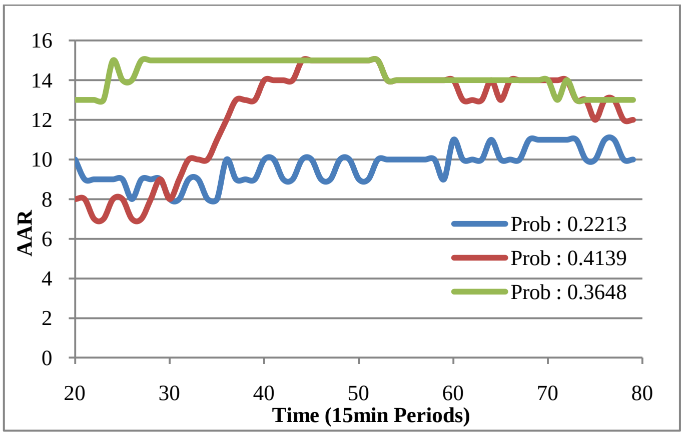

Figure 5. Probabilistic Profiles for cluster 2

We call this procedure ***TAF Clustering***

E. Dynamic Time Warping Profiles

Dynamic Time Warping (DTW) is an established methodology to study the similarity between two electrical signals. It has been used in the areas of speech recognition to match speech patterns. Reference [14] has demonstrated that DTW can be useful to detect similar multidimensional time series. DTW is a technique where one sequence is "warped" in time around the other. A distance matrix is generated between all the time pairs of the two series. A least expensive path in the distance matrix is determined dynamic programming. The path is subjected to certain end point constraints. We use the technique of dynamic time warping to compare the day of operations TAF with historic TAFs. AAR capacities of the historically similar TAF days are the actual capacity profiles and the probabilities of the profiles are proportional to the least expensive path length between the historically similar TAF with the day of operations TAF.

The multidimensional time series is the TAF for the day-of-operation and we have used the Euclidean norm to generate the distance matrix. Further we have restricted the search of the least expensive path to 3 periods as we want to search for similarity within a neighborhood of a period. Once we have the least expensive path lengths between the day of operations TAF and all the historic TAF days we can find the days which are most similar to the day of operation TAF. The mathematical formulation takes the following form.

Let T_d be a time series of the TAF for the day of operation. T_d is thus a 7 dimensional time series of length 60. The 7 dimensions represent the forecast in a period and the day-of-operation is divided in 60 periods.

Let $\{T_i\}_{i=1}^{N}$ be a set of TAFs. DTW evaluates the least expensive path between T_i and T_j. A distance matrix of size 60 × 60 is first computed. Any element (r,s) of the distance matrix is $D(r,s) = \|T_i(r) - T_j(s)\|_2^2$ (r,s $\in [1..60]$). $T_i(r)$ is the TAF for period *r* for day *i* and similarly $T_j(s)$ is the TAF for period *s* for day *j*. Thus a distance matrix for all possible day pairs is computed i.e. a total of ${}^N_2C$ matrices are computed.

The DTW path between T_i and T_j is thus

$$DTW(T_i, T_j) = D[T_i(t), T_j(t)] + \min\{DTW[T_i(t-1), T_j(T-1)], DTW[T_i(t-1), T_j(t)], DTW[T_i(t), T_j(t-1)]\} \quad (13)$$

This path has to start from t=1 and has to end at t=60. These are the end points conditions.

So if two sequences, A and B, are completely similar i.e. identical, the DTW(A,B) is 0. The optimal path would be the diagonal of the distance matrix.

In actual operations the day-of-operation TAF should be compared to the TAF of the days which precede it but due to data limitations we considered all the days. We experimented with the number of profiles following rule to determine the number of profiles:

Number of profiles for the day i with TAF T_i

$$n = argmax_k \left[\sum_{j=1}^{k} DTW(T_i, \{T_{[j]}\}) \leq Const \right] \quad (14)$$

Const is a tuning parameter determined by experimentation with the number of scenarios. We found our solution to be relatively stable for the profiles between 8 to 12 and based on this adopted ***Const*** = 2500. A lower value of ***Const*** would reduce ***n***. If the TAF is accurate uncertainty in the AAR can represented by a few profiles. If TAF is inaccurate then less scenarios would imply that we may be ignoring certain representations of the AAR which would increase the cost of delays.

Where {T[j]} is an ordered set such that

$$DTW(T_i,T[1]) \leq DTW[T_i,T[2])] \leq DTW[T_i,T[3])] \leq .. \leq DTW(T_i,T[N]) \quad (15)$$

The set of profiles is thus the actual AARs for the 'n' days.

$$S[k] = AAR[k] \quad (\forall k \in ([1],..,[n])$$

The profile probabilities are obtained after normalizing the least resistant path for the 'n' days.

$$P[k]=\frac{1-\left(DTW(T_i,\{T_{[k]}\}) \Big/ \sum_{j=1}^{n} DTW(T_i,\{T_{[j]}\})\right)}{n-1} \quad (16)$$

We call this procedure as ***DTW Profiles***.

IV. COMPARATIVE COMPARISONS

This section compares the cost of the delays after implementing the probabilistic capacity profiles developed in Section III in an ATFM simulation. The ATFM simulation is conducted using the GDP model developed in [3]. This model minimizes the total of cost of delay in a GDP by determining the arrival rate at the airport and the number of the aircraft subjected to ground delays. The model outputs the number of planes which can land in the absence of air holdings, the Planned Airport Arrival Rate (PAAR). As mentioned in the introduction, uncertainty of the AAR is captured by the probabilistic capacity profiles. In the model the cost of air delay c_a is assumed to be greater than c_g (otherwise there would not be a need to ground hold the aircraft). The model takes the following form:

$$Min\left[\sum_{t=1}^{T} c_g \times G(t) + \sum_{p=1}^{N}\sum_{t=1}^{T} c_a \times W(S_p,t) \times P_p\right] \quad (17)$$

Subject to:

$$A(t) - G(t-1) + G(t) = D(t)$$
$$\{\forall t \in 1..T+1, G(0) = G(T+1) = 0\} \quad (18)$$

$$-W(S_p,t-1) \pm W(S_p,t) - A(t) \geq -M(S_p,t)$$
$$\begin{Bmatrix}\forall t \in 1..T+1, -W(S_p,0) = -W(S_p,T+1) = 0, \\ p \in 1..N\end{Bmatrix} \quad (19)$$

$$A(t), W(S_p,t), G(t) \in Z_+ \{\forall t \in 1..T+1, p \in 1..N\} \quad (20)$$

Where, **t**: is the time period, $\mathbf{S_p}$: p^{th} capacity profile (length T)**;** P_p: is the probability of profile ***p*****;** **T**: total number of time periods or planning horizon**;** **N**: total number of profiles**;** **G(t)**: ground holding at time **t;** $\mathbf{W(S_p,t)}$: air holding under profile S_p at time t;**A(t)**: planned airport acceptance rate at time t **(PAARs);** $\mathbf{M(S_p,t)}$: capacity under profile $\mathbf{S_p}$ at time t**;** **D(t)**: demand in period t**;** $\mathbf{c_a}$: cost of airborne delay**;** $\mathbf{c_g}$: cost of ground delay**;** $\{S_p\}_{p=1}^{N}$ is the set of profiles;$\sum_{p=1}^{N} P_p = 1$

The objective function, (17), minimizes the sum of the fixed ground delay costs and the (expected) air delay costs. Equation (18), is a queuing constraint for flights bound for the destination from all the origin airports. It enforces flow conservation. The demand at period t, ***D(t)***, plus the backlogged planes ground held in period t-1,***G(t-1)***, should either land, ***A(t)***, or be put in a queue, ***G(t)***. Equation (19) is a queuing constraint at the destination airport. Under capacity profile ***p***, all planes that can land, ***A(t)***, or air delayed from the previous time period $W(S_p,t-1)$ either land or are further air delayed to the next period, $W(S_p,t)$. The inequality is required as the demand might be less than the available capacity (runway + airspace). Equation (20) ensures that ***A(t)***, $W(S_p,t)$ and ***G(t)*** are real positive integers.

The decision variables are the number of aircrafts landing in a period t, ***A(t)***, the number of aircrafts which are subjected to ground holding ***G(t)*** and the number of aircrafts subjected to air borne holding under profile ***p***, $W(S_p,t)$. The ratio of the cost of delays is selected to be 3:1::c_a:c_g based on heuristics and established conventions (the cost of air delays is roughly 3 times the cost of ground delay c_a is expensive due to higher fuel consumption in air, crew costs etc). The data for demand, ***D(t)*** (planes originally scheduled to land in a period t) is obtained from the ASPM website. The model was solved in AMPL using the CPLEX as the solver with a run time of less than a second.

We simulated ATFM strategies for 50 historical days from 2004 to 2006 when the low lying marine stratus was observed at SFO and the results are based on these days. We generated the stochastic capacity profiles for each historical day using their TAF and STRATUS forecasts. For the Naïve case the profiles and probabilities were the same across all the days as the profiles are generated without weather forecast information. In case of the TAF Clustering, we first determined in which of the two clusters, c_1 or c_2, the historical days TAF belong to and then applied the profiles and probabilities under that particular TAF cluster in the GDP model. So the uncertainty in the historical days AAR is represented by either 2 or 3 probabilistic capacity profiles depending on the classification of its TAF. Using the GDP model, we were able to compare the benefits under the different forecasts and different methodologies.

The GDP model determines four PAARs for the any historical day: $PAARs_{fog}$, $PAARs_{TAF}$, $PAARs_{Naive}$ and $PAARs_{DTW}$. The PAARs are obtained using the probabilistic capacity profiles generated from Fog Clustering, TAF Clustering, Naïve Clustering and DTW Profiles methodologies respectively. To understand if the forecasts are useful in controlling the costs of delays, the four sets of PAARs are compared in a deterministic queuing model. The idea for the deterministic queuing model is simple: the PAAR for a period, maybe higher than the actual realized capacity resulting in airborne delays from a capacity-PAAR imbalance. This also measures the similarity of the profile to the actual realized capacity. There are two types of delays if the PAAR is implemented 1) Airborne delays between the PAAR and the actual realized capacity 2) Ground delays between the PAAR and original schedule. The total cost (TC) of delay is $c_g \times$ *ground delay* + $c_a \times$ *airborne delay*. We obtain TC_{fog}, TC_{naive}, TC_{TAF} and TC_{DTW} from the queuing model.

The various total costs of delay are compared to a ***Perfect Information*** (PI) case where the controllers have perfect foresight about the evolution of capacity as if told by an "oracle". For any historical day, we know the actual realized capacity and this capacity can be used in a (deterministic) ATFM simulation. This is equivalent to having one profile which is the actual AAR profile with 100% probability of occurrence in the GDP model. Thus if we have perfect information we can subject ground holding to all the aircrafts which would experience air borne delays under the original schedule. We translate all the possible air borne delays to ground delays, lowering total costs. Therefore, TC_{PI} = ground delay. TC_{PI} is the minimum possible cost that can achieved for the day-of-operation. The average costs of delay for 50 days, obtained after the deterministic queuing model are shown in Table II.

From the average cost of delays it can be seen that average TC_{PI}< average TC_{fog} <average TC_{DTW} < average TC_{TAF}<average TC_{naive}. This illuminates the fact that probabilistic profiles derived from weather forecasts are better in planning of operations as compared to profiles developed devoid of any forecast information. The STRATUS forecast gives the minimum average cost of delays. This cost is marginally higher than the average delay from the DTW Profiles. The average costs of both STRATUS and DTW are almost double the average costs of PI. Days which have lower DTW costs than Fog clustering indicate that the TAF is accurate in predicting actual weather where as STRATUS is not accurate on those days. Days when Naïve clustering gives a lower cost than DTW and Fog clustering indicate that both the forecasts were inaccurate. The costs of delays from the probabilistic capacity scenarios rely on the quality of the forecast, if the forecast is accurate in predicting the capacity, the delays would be minimized. The results indicate that TAF can be used to integrate weather with ATFM decision making for all airports.

V. CONCLUSIONS

In this paper we have demonstrated how to integrate the weather forecasts to generate probabilistic scenarios for the day of the operation. We concentrated on the summer months (May-September) from 2004 to 2006. This is the first step towards the incorporation of weather forecast information in NEXTGEN Though, we have considered SFO in our experiments, the methodologies can be applied to any airport. It is shown that incorporating the weather information to plan the day of operation arrivals leads to reduced costs. It is important to note that STRATUS is designed specifically for SFO and particularly for the days when there is a low lying stratus over the airport thus its application for all airports is limited. From our cost of delay calculations it can be seen that the difference of the costs between DTW Profiles and Fog Clustering is small, so the TAF offers similar benefit in decision making when incorporating forecast information to construct the probabilistic capacity profiles. Thus the DTW methodology can be applied to all airports.

REFERENCES

[1] A.Mukherjee and M. Hansen, "A Dynamic Stochastic Model for the Single Airport Ground Holding Problem," Transportation Science Vol. 41, No. 4, November 2007, pp. 444-456

[2] David A. Clark, "Investigating a New Ground Delay Program Strategy for coping with SFO Stratus," 89th AMS Annual Meeting ARAM Special Symposium on Weather - Air Traffic Phoenix, AZ, 11-15 January 2009

[3] M. Ball O., R. Hoffman, A. R. Odoni and R. Rifkin, " A Stochastic Integer Program with dual network structure and its application to the Ground-Holding Problem,". Oper. Res. 51 167–171. 2003

[4] O Richetta and A. R. Odoni, " Solving optimally the Static Ground Holding Policy Problem in air traffic control," Transportation Science 27, pp 228–238, 1993

[5] P. Liu, M. Hansen and A. Mukherjee, "Scenario-based air traffic flow management: From theory to practice," Transportation Research Part B: Methodological, Volume 42, Issues 7-8, August 2008, Pages 685-702

[6] M. Kaut and S.W Wallace, "Evaluation of Scenario-generation Methods for Stochastic Programming," Stochastic Programming E-Print Series, 2003-14, <http://hera.rz.hu-berlin.de/speps/>.

[7] L. Cook and B. Wood, "A Model for Determining Ground Delay Program Parameters using a probabilistic forecast of Stratus clearing," Eighth USA/Europe Air Traffic Management R&D Seminar (ATM2009), Napa, United States 2009

[8] A. Klein, R. Jehlen and D. Liang, "Weather Index With Queuing Component For National Airspace System Performance Assessment," 7th USA/Europe ATM R&D Seminar(ATM2007), Barcelona, Spain, 2007.

[9] SAS Institute, 2004 SAS Institute, 2004. SAS/STAT 9.1 User's Guide. Cary, NC.

[10] L. Kaufman and P. Rousseeuw, Finding Groups in Data: An introduction to Cluster Analysis, Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics, New York: Wiley, 1990

[11] D.A. Clark et al., SFO Marine Stratus Forecast System Documentation, Project Report ATC-319, MIT Lincoln Laboratory, Lexington, MA, 2006.

[12] I. Jolliffe, Principal Component Analysis,Springer Verlag, New York, 1986.

[13] C. Ding , X. He, "*K*-means clustering via principal component analysis,", Proceedings of the twenty-first international conference on Machine learning, p.29, July 04-08, 2004, Banff, Alberta, Canada

[14] M. Vlachos, M. Hadjieleftheriou, D. Gunopulos and E. Keogh, "Indexing Multi-Dimensional Time-Series with Support for Multiple Distance Measures",In Proc. of 9th International Conf. on Knowledge Discovery & Data Mining (SIGKDD), Washington, DC, 2003

TABLE II. AVERAGE COST OF DELAYS

Method	PI	DTW	TAF Clustering	Naïve	Stratus
Average	96.5	194.17391	214.46	239.17	182
Std. Dev	54.9	125.89802	145.28	156.72	109.93