Ninth USA/Europe Air Traffic Management Research and Development Seminar (ATM2011)

# Generating Probabilistic Capacity Profiles from weather forecast: A design-of-experiment approach

Gurkaran Buxi
National Centre of Excellence for Aviation Operations Research
NEXTOR, University of California Berkeley
Berkeley, CA
gkb@berkeley.edu

Mark Hansen
National Centre of Excellence for Aviation Operations Research
NEXTOR, University of California Berkeley
Berkeley, CA
mhansen@ce.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 develops probabilistic profiles for three airports, BOS, LAX and SFO using the Terminal Aerodrome Forecast and San Francisco Marine Initiative. The profiles are inputs to a static stochastic GDP model to simulate ATFM strategies. A design of experiments approach has been employed to determine profiles which minimize the total average costs. The average cost of the methodologies is evaluated against realized capacities to determine the benefit of the forecast. It is shown that inclusion of weather forecasts reduces the cost of delays. It is shown for SFO that on average TAF offers similar benefit in controlling cost of delay when compared to STRATUS. Careful use of the TAF indicates that other airports would also benefit from using TAF in planning of operations**.

***Keywords-ground delay program;terminal aerodrome forcast ;STRATUS;Design of experiments;Response surface methodology;Dynamic time warping***

## I. INTRODUCTION

Adverse weather conditions in the vicinity of an airport often reduce its operational capacity, leading to an imbalance between capacity and demand. This capacity-demand imbalance may lead to delays, and, in the absence of traffic management initiatives, holding in the terminal area, increased controller workload, and excessive fuel burn. To mitigate these impacts, the Federal Aviation Administration (FAA) often implements Ground Delay Programs (GDPs). GDPs mitigate the terminal 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 arrival capacity forecast or 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. There has been considerable research on how to plan GDPs so as to take into account uncertainty about airport capacity. 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 information about realized capacity is updated, ground holding decisions are revised, incorporating a *wait-and-see* strategy. Most 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. Most of the literature on these models has taken the capacity profiles or scenario trees as given, assuming that in real-world application these could somehow be extracted from weather forecasts and the expertise of traffic management specialists. There is considerably less literature on the development of specific day-of-operation probabilistic capacity profiles. Accordingly, this paper focuses on the development of probabilistic capacity profiles from a day-of-operation weather forecast using a design-of-experiments (DOE) methodology. This methodology determines the best input parameter values which lower the costs in a GDP. Such profiles, when used in conjunction with appropriate GDP planning models, could lead to better GDPs, with lower realized costs as a result of reducing either excessive ground delays or airborne delay.

This paper develops probabilistic capacity profiles from weather forecasts for three United States (US) airports: San Francisco International Airport (SFO), Boston Logan

International Airport (BOS) and Los Angeles International Airport (LAX). The weather forecast used for constructing the profiles for the three airports is the Terminal Aerodrome Forecast (TAF) which is issued for all the major US Airport. TAF contains forecast information on visibility, ceiling, winds, and other meteorological variables for the entire day. Amongst the above airports SFO is unique because it is issued another forecast, SFO Marine Stratus Forecast System (STRATUS) along with the TAF. STRATUS is a forecast project created specifically for SFO, because it 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 the TAF for all the three airports and also construct profiles for SFO using the STRATUS forecast.

The contribution of this paper is that it provides techniques which use several statistical methodologies to convert weather forecasts into specific day-of-operation probabilistic capacity profiles using a design-of-experiments (DOE) approach. The profiles are provided as inputs to a static stochastic GDP model that determines the optimal arrival rate for an airport. The DOE approach determines parameters which generate probabilistic profiles minimizing the total realized costs. We compare the realized costs of the simulated outcomes from the GDP model based on the different methods of scenario generation from weather forecasts against two other reference cases. In the first, the GDP is based on perfect information about the capacity, while in the second profiles are developed from historical capacity data without use of the day-of-operation weather forecast.

This paper develops probabilistic capacity profiles based on the realized capacity and the weather forecast for the summer months (May-September) of 2004 to 2006 for the three airports. In total, the data set included D=446 days for SFO, D=432 days for BOS and D=450 days for LAX for which the TAF weather forecast and the realized capacity were both available. The STRATUS forecast for SFO was available for only 150 days because they become available when marine clouds are forecast in the in the terminal area. We construct probabilistic profiles which represents capacity for every 15 minutes (period) from 7am to 10pm as the bulk of the traffic is occurs in this time period. The reported results are based on three airports for 45 historical days but the methods for generating capacity profiles from the TAF can be applied at any other airport for which a TAF is available.

This paper proceeds as follows. Section 2 provides the literature review. Section 3 presents the GDP model Section 4 describes the weather forecasts and the several techniques for generating probabilistic capacity profiles using the design of experiment approach. Section 5 presents a cost comparison of the strategies obtained from the profiles developed in Section 4. Section 6 offers conclusions.

## II. LITERATURE REVIEW

The current National Airspace System (NAS) rarely incorporates uncertainty of the weather forecasts into strategic 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 probabilistic capacity profiles or scenario trees for AAR as inputs [1], [3], [4]. A probabilistic capacity profile is a time series of capacity values (typically based on a quarter-hour time unit) and an associated probability. For a given airport and day there will typically be several profiles depicting different possible evolutions of capacity. Thus the set of stochastic profiles capture the uncertainty in the future arrival capacity. Methods for generating these profiles have focused on developing them from historical data without specific reference to a particular day-of-operation [5]. Other scenario-generation methods have been developed to support the application of stochastic programs in finance [6].

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] presents a GDP model based on the SFO Marine Stratus Initiative (STRATUS) forecast. They model the time of fog burn off as a random variable with the probability distribution obtained from STRATUS. They assume at fog burn-off the landing capacity of SFO increases sharply. Refence [8] gauged the imprecision of the forecast weather information with the actual weather by calculating avoidable delays. First they matched the realized historical weather in a period with the capacity of the airport in that period. Using this developed relationship, they predicted the AAR from the Terminal Aerodrome Forecast (TAF) and the Meteorological Aviation Report (METAR) for every period. From a queuing model, they determined the delays between the scheduled arrivals and AAR predicted from TAF and AAR predicted from METAR. This deterministic approach ignores the uncertainty concerning the TAF. Reference [9] uses the day-of-operation weather forecast to generate probabilistic capacity profiles for SFO but have not addressed how to choose the parameters which influence the probabilities and the number of the profiles. Their approach does not produce probabilistic profiles which give the lowest realized costs in a GDP simulation.

While there is previous research that addresses the development probabilistic capacity profiles from historical capacity data and which translates a TAF forecast into a deterministic capacity forecast, the problem of developing probabilistic capacity scenarios from a TAF forecast that give the lowest costs has yet to be addressed. The research presented here fills that gap.

## III. BALL ET AL. STATIC STOCHASTIC GDP MODEL

This section describes Ball et al. [3] static stochastic model which requires probabilistic capacity profiles as inputs. This model determines ground delays by minimizing the total expected costs of delay in a GDP by determining the optimal rate at which aircraft should land at the destination airport. This rate is termed as the Planned Airport Arrival Rate (PAAR), for each time period. As mentioned in the introduction, uncertainty of the AAR is captured by probabilistic capacity profiles. In the model the cost of air delay $c_a$ is assumed to be greater than cost of ground delay $c_g$ (if $c_a \le c_g$ there would not be a need to delay the aircraft on the ground). The model takes the following form:

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

Subject to:

$$\boldsymbol{A(t) - G(t-1) + G(t) = D(t)}$$
$$\{\forall \boldsymbol{t} \in \boldsymbol{1..T+1}, \boldsymbol{G(0) = G(T+1) = 0}\} \tag{2}$$

$$\boldsymbol{-W(S_p, t-1) + W(S_p, t) - A(t) \ge -M(S_p, t)}$$
$$\left\{\begin{matrix}\forall \boldsymbol{t} \in \boldsymbol{1..T+1}, \boldsymbol{-W(S_p, 0) = -W(S_p, T+1) = 0}, \\ \boldsymbol{p} \in \boldsymbol{1..N}\end{matrix}\right\} \tag{3}$$

$$\boldsymbol{A(t), W(S_p, t), G(t)} \in \boldsymbol{Z_+}\ \{\forall \boldsymbol{t} \in \boldsymbol{1..T+1}, \boldsymbol{p} \in \boldsymbol{1..N}\} \tag{4}$$

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

The objective function, (1) minimizes the sum of the fixed ground delay costs and the expected air delay costs. Equation (2), 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 planes ground held in period t-1,***G(t-1)***, should either land, and thus count toward ***A(t)***, or be put in a queue, contributing to ***G(t)***. Equation (3) is a queuing constraint at the destination airport. Under capacity profile $\boldsymbol{S_p}$, all planes can land, ***A(t)***, or are 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. Equation (4) 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 holding under profile $\boldsymbol{S_p}$, ***W($S_p$,t)***. The ratio of the cost of delays is selected to be $c_a/c_g$=3 based on published data. The data for demand, ***D(t)*** (planes originally scheduled to land in a period t) is obtained from the ASPM website.

This GDP model determines a PAAR for a given demand profile and a set of probabilistic capacity profiles. According to this model, the aircraft arrive, intending to land at the destination airport, at the rate determined by the PAAR. When the aircraft approach the terminal there might insufficient capacity as the inclement weather might have persisted. This *realized* capacity might be different than the capacity represented by the probabilistic capacity profiles. This leads to an additional *realized* airborne delay at the airport. If shown by a queuing diagram, this realized airborne delay is the area between the PAAR and the realized capacity curves. The difference between the PAAR and the realized capacity at any time determines the amount of realized airborne holding, while the ground holding is obtained directly from the GDP model. The total realized cost (TC) of delay for any day-of-operation is $c_g \times$ *ground delay* + $c_a \times$ *airborne delay. The extent of the benefit of weather forecasts in decision making is gauged by comparing the realized TC averaged over a sample of historical days. The average TC under different methods of profile generation is also compared.*

In the next section, we provide several methodologies to generate the profiles ($\boldsymbol{S_p}$), their probabilities ($\boldsymbol{P_p}$) and the number of profiles (***N***) required as inputs by the Ball et al model.

## IV. Profile Generation and weather forecasts

In this section we first discuss the methodology to generate profiles devoid of weather forecast information. We then discuss the Terminal Aerodrome Forecast (TAF) and the methodologies to generate probabilistic capacity profiles from the TAF. We conclude the section by discussing the STRATUS forecast and the associated methodology for generating profiles exclusively for SFO from STRATUS. We have collected historical realized capacity, the TAF and the scheduled demand for N historical days.

### *A. No Weather Forecast: Naïve Clustering*

This method of profile generation does not incorporate any weather forecast and is similar to that described in [5]. The methodology generates profiles from realized historical capacity and these are used as probabilistic profiles for the day-of-operation. In [5], the centroids of the clusters obtained after K-means clustering on the AAR time series are the profiles. Let $[\mathrm{A}]_{\mathrm{T\times N}} = [\mathrm{A}_1, \mathrm{A}_2, .., \mathrm{A}_\mathrm{N}]$ be the data matrix of the AAR time series for D historical days. $\mathbf{A_r}$ is column vector of the AAR time series of length T time periods for day r. K-means clustering splits the data matrix into a predefined number of clusters, ***l***, where each cluster $c_k$ contains $\boldsymbol{d_k}$ days. The days which have similar AAR time series are grouped together i.e. they are in the same cluster. The similarity is defined as the sum over all the periods of the Euclidean norm between the AAR time series for each period. A smaller value of the Euclidean norm implies greater similarity. After the K-means operation we obtain a partition of a set of days.

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

Such that,

$$\sum_{j=1}^{l} \boldsymbol{d_j} = \boldsymbol{N} \quad \text{and} \quad \bigcup_{j=1}^{l}\{\boldsymbol{A_h^j}\}_{\boldsymbol{h=1}}^{\boldsymbol{d_j}} = \{\boldsymbol{A_h}\}_{\boldsymbol{h=1}}^{\boldsymbol{N}} \quad \text{and} \quad \bigcap_{j=1}^{l}\{\boldsymbol{A_h^j}\}_{\boldsymbol{h=1}}^{\boldsymbol{d_j}} = \boldsymbol{\Phi} \tag{5}$$

The optimal number of clusters $l^*$, is an open problem and there are ad-hoc procedures which assist in determining it. More clusters imply more profiles which capture more variation in capacity but each profile then has a lower probability of occurrence. Reference [5] provides an algorithm to determine the number of clusters involving the pseudo- F statistic while enforcing a minimum number of days for ($d_k$) for each cluster.

Procedures like the pseudo- F statistic 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 [10]. 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. An alternative is the Silhouette value, which varies 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 [11]. 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, it might be worth 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 $S_i$ is determined by the average of the AAR profiles in the cluster $c_i$.

$$\boldsymbol{S_i} = \left\lceil \sum_{h=1}^{d_i} \boldsymbol{A_h^i} \Big/ \boldsymbol{d_i} \right\rceil \quad \boldsymbol{i \in 1..l^*} \tag{6}$$

Where [ ] is the nearest integer roundup operator.

The probability of the profile $S_i$ is the proportion of days in $c_i$. $\boldsymbol{P_i = d_i / N} \quad \boldsymbol{i \in 1..l^*}$. The number of clusters was determined by the highest average silhouette value.

We call this procedure ***Naïve Clustering*** as it clusters the AAR without any weather information.

*B. TAF-based Capacity Profiles*

The Terminal Aerodrome Forecast (TAF) is a weather forecast issued for every major airport four times a day at 6hour intervals by the National Oceanic and Atmospheric Administration (NOAA). It contains numeric values for seven metrological variables (wind speed and direction, visibility, and heights of four cloud types) along with qualitative variables for rain, fog, mist, etc for that airport. The TAF issued between 5am and 7am was used for developing *two* methodologies of generating probabilistic capacity profiles. Since, the TAF forecasts seven metrological variables for each period and therefore the entire day-of-operation can be represented by a column vector of length 60 (15 min periods) ×7 variables/period = 420. Therefore, the entire TAF data set could be represented by a 420 (variables) × N( total number of historical days) matrix.

Let $[\mathbf{T}]_{\mathbf{L\times D}}$ be a matrix (L=420, N= total number of days), 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, representing 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 (7) through (11) describe the PCA on the data set.

$$[\mathbf{C}]_{\mathbf{L\times L}} = [\mathbf{T}][\mathbf{T}]^{\mathbf{t}} \big/ \mathbf{N-1} \tag{7}$$

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

$$\mathbf{C}\boldsymbol{X} = \boldsymbol{\lambda X} \tag{8}$$

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

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

$$\boldsymbol{\lambda_{[1]} > \lambda_{[2]} > \lambda_{[3]} > .\,. > \lambda_{[L]}} \tag{9}$$

A standard technique is to capture 90% variability, in which case the number of eigenvalues required is given by Equation 10.

$$\boldsymbol{n = argmax_k \frac{\sum_{p=1}^{k} \lambda_{[p]}}{\sum_{p=1}^{L} \lambda_{[p]}} \leq 0.9} \tag{10}$$

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

The reduced TAF matrix

$$[\widetilde{\mathbf{T}}]_{\mathbf{n\times D}} = [\mathbf{W}] \times [\mathbf{T}] \tag{11}$$

In this reduced TAF matrix, $[\widetilde{\mathrm{T}}]_{\mathrm{n\times N}}$ we wanted to classify days which had the similar TAF. 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.

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

$$\left\{\widetilde{\boldsymbol{T}}_{\boldsymbol{h}}^{\boldsymbol{1}}\right\}_{\boldsymbol{h=1}}^{\boldsymbol{d_1}}, \left\{\widetilde{\boldsymbol{T}}_{\boldsymbol{h}}^{\boldsymbol{2}}\right\}_{\boldsymbol{h=1}}^{\boldsymbol{d_2}}, \left\{\widetilde{\boldsymbol{T}}_{\boldsymbol{h}}^{\boldsymbol{3}}\right\}_{\boldsymbol{h=1}}^{\boldsymbol{d_3}}, \ldots, \left\{\widetilde{\boldsymbol{T}}_{\boldsymbol{h}}^{\boldsymbol{l}}\right\}_{\boldsymbol{h=1}}^{\boldsymbol{d_l}} \tag{12}$$

Such that,

$$\boldsymbol{\sum_{j=1}^{l} d_j = N} \tag{13}$$

Where $d_j$ is the number of days in the cluster $c_j$

$$\cup_{\boldsymbol{j=1}}^{\boldsymbol{l}} \left\{\widetilde{\boldsymbol{T}}_{\boldsymbol{h}}^{\boldsymbol{j}}\right\}_{\boldsymbol{h=1}}^{\boldsymbol{d_j}} = \left\{\widetilde{\boldsymbol{T}}_{\boldsymbol{h}}\right\}_{\boldsymbol{h=1}}^{\boldsymbol{N}} \text{ and } \cap_{\boldsymbol{j=1}}^{\boldsymbol{l}} \left\{\widetilde{\boldsymbol{T}}_{\boldsymbol{h}}^{\boldsymbol{j}}\right\}_{\boldsymbol{h=1}}^{\boldsymbol{d_j}} = \boldsymbol{\Phi} \tag{14}$$

After the PCA operation, the variables are uncorrelated and thus the number of clusters is determined using the pseudo-F statistic. Let $\boldsymbol{l^*}$ be the optimal number of TAF clusters, thus from this analysis on the TAF we can classify the day-of-operation in either one of the $\boldsymbol{l}^*$ clusters from $\boldsymbol{c_1}$ to $\boldsymbol{c_{l^*}}$ depending on the classification of its TAF.

Next, we determined a set of representative capacity profiles from the realized capacity of the days which were classified in the same TAF cluster. We performed another K-means clustering on the realized AAR time series profile of the days within $\boldsymbol{c_1}$ to $\boldsymbol{c_{l^*}}$. Let $\{\widetilde{\mathrm{A}}_{\mathrm{h}}^{\mathrm{r}}\}_{\mathrm{h}=1}^{\mathrm{d_r}}$ be the set of AAR time series for the days in cluster $\boldsymbol{c_r}$ ($\mathrm{r} \in 1..l^*$). The K-means operations partitioned the AAR data set within each TAF cluster. The highest average Silhouette value determined the number of clusters in the second K-means clustering. Let the optimal number of AAR clusters within a TAF cluster $\boldsymbol{c_r}$ be $\boldsymbol{k_r^*}$ ($\mathrm{r} \in 1..\boldsymbol{l}^*$). Define $\{\widetilde{\mathrm{A}}_{\mathrm{h}}^{\mathrm{r,i}}\}_{\mathrm{h}=1}^{\mathrm{d_{r,i}}}$ to be the set of AAR profiles for days in an AAR cluster i within a TAF cluster $\mathrm{c_r}$ and $\mathrm{d_{r,i}}$ are the total number of days within AAR cluster i and TAF cluster $\boldsymbol{c_r}$ ($\mathrm{i} \in 1..\boldsymbol{k_r^*}$, $\mathrm{r} \in 1..\boldsymbol{l}^*$).

$$\bigcup_{i=1}^{k_r^*}\{\widetilde{\mathrm{A}}_{\mathrm{h}}^{r,i}\}_{\mathrm{h}=1}^{\mathrm{d_{r,i}}} = \{\widetilde{\mathrm{A}}_{\mathrm{h}}^{r}\}_{\mathrm{h}=1}^{\mathrm{d_r}} \text{ and } \sum_{i=1}^{k_r^*} \boldsymbol{d_{r,i}} = \boldsymbol{d_r} \tag{15}$$

The set of capacity profiles are the averages of the AAR time series in the AAR cluster. The probability of the profile is the proportion of days within the AAR cluster

The profiles and their probability are determined as below

$$\boldsymbol{S_{r,i}} = \left[ \sum_{h=1}^{d_{r,i}} \widetilde{A}_h^{r,i} \Big/ d_{r,i} \right] \tag{16.1}$$

$$\boldsymbol{P_{r,i}} = \boldsymbol{d_{r,i}} \Big/ \sum_{k=1}^{k_r^*} \boldsymbol{d_{r,k}} \quad \boldsymbol{i} \in \mathbf{1}..\boldsymbol{k_r^*}, \boldsymbol{r} \in \mathbf{1}..\boldsymbol{l}^* \tag{16.2}$$

We call this procedure ***TAF based clustering***.

## C. Dynamic Time Warping Profiles

This is the second method which uses the TAF to determine the probabilistic capacity profiles using Dynamic Time Warping (DTW). DTW is an established methodology to study the similarity between two electrical signals. It is particularly useful to match sequences which are translated in time. Recent research has demonstrated that DTW can be useful to detect similar multidimensional time series [14]. DTW is a technique where one sequence is "warped" in time around the other. The two time series are aligned to a distance matrix such that both of them start from the lower left corner and end at the top right corner. Each cell of the distance matrix is a cost representing the distance between the corresponding time pairs of the two series. Finally, a minimum cost path between the lower left corner and the upper right corner of the distance matrix is determined using dynamic programming. The cost of this path represents the similarity between the time series. The deviation of the minimum cost path from the diagonal of the distance matrix indicates the warping of the time series. In this case, the multidimensional time series being compared are the day-of-operation TAF and the historical TAFs. DTW can match TAF variables from different time periods and therefore can determine forecasts which are similar but translated in time. We used the Euclidean norm, applied to the standardized weather variables, to generate the costs for each cell in the distance matrix.

Potentially, several minimum cost paths are possible through the distance matrix. To restrict the paths, we have multiplied the off-diagonal cells by a Weighing Factor (WF≥ 1). A higher WF restricts the warping of the time series and aligns the minimum cost path closer to the diagonal of the distance matrix whereas a lower WF allows the minimum cost path to vary through the distance matrix. In other words a high WF compares forecasts locally and not for the entire day. Tables I and II, shows the distance matrix and minimum cost path (highlighted) along with the cost for two artificially generated multidimensional time series of length 5 periods.

TABLE I. MINIMUM COST PATH FOR WF=1

| **WF = 1, Minimum cost path = 0.962** | | | | | |
|---|---|---|---|---|---|
| ***Periods*** | ***1*** | ***2*** | ***3*** | ***4*** | ***5*** |
| ***5*** | 0.498 | 0.412 | 0.620 | 0.116 | 0.093 |
| ***4*** | 0.815 | 0.376 | 0.943 | 0.184 | 0.048 |
| ***3*** | 0.200 | 0.688 | 0.532 | 0.117 | 0.648 |
| ***2*** | 0.926 | 0.119 | 0.448 | 0.251 | 0.560 |
| ***1*** | 0.136 | 0.815 | 0.814 | 0.062 | 0.422 |

TABLE II. MINIMUM COST PATH FOR WF=10

| **WF = 10, Minimum cost path =1.07** | | | | | |
|---|---|---|---|---|---|
| ***Periods*** | ***1*** | ***2*** | ***3*** | ***4*** | ***5*** |
| ***5*** | 4.978 | 4.122 | 6.197 | 1.160 | 0.093 |
| ***4*** | 8.149 | 3.760 | 9.433 | 0.184 | 0.482 |
| ***3*** | 1.999 | 6.879 | 0.532 | 1.175 | 6.481 |
| ***2*** | 9.264 | 0.119 | 4.483 | 2.507 | 5.600 |
| ***1*** | 0.136 | 8.152 | 8.138 | 0.618 | 4.225 |

We use the technique of dynamic time warping to compare the day-of-operations TAF with historic TAFs. AAR time series of the historically similar TAF days are used for the capacity profiles. The probabilities of the profiles are inversely proportional to the cost of the minimum cost path raised to a Dimension Factor (DF). As the DF increases, the degree of similarity decreases because the days which are less similar to the day-of-operation are penalized greater by an increase in their total cost.

The mathematical formulation takes the following form. Let $\mathrm{F_D}$ be a time series of the TAF for the day-of-operation. $\mathrm{F_D}$ is thus a 7 dimensional time series of length 60. The 7 dimensions represent the forecast per quarter period and the day-of- operation is divided in a total of 60, 15 minute periods.

Let $\{\mathrm{F_j}\}_{\mathrm{j}=1}^{\mathrm{N}}$ be a set of N historical TAFs for N historical days. DTW evaluates the minimum cost path between $\mathrm{F_D}$ and $F_j$. $\mathrm{F_D}(\mathrm{r})$ is the TAF for period r for the day-of-operation and similarly $\mathrm{F_j}(\mathrm{s})$ is the historical TAF for period s for day j. A distance matrix of size 60 × 60 is first computed for all possible pairs i.e. a total of N matrices are computed for comparison. Any element (r,s) of the distance matrix for comparing $F_D$ and

$F_j$ is $D(F_D(r), F_j(s)) = \|F_D(r) - F_j(s)\|_2^2 \times WF$ $(r,s \in [1..60], r \neq s)$ and $D(F_D(r), F_j(r)) = \|F_D(r) - F_j(r)\|_2^2$ $(r \in [1..60])$

The minimum cost path between $F_D$ and $F_j$ is given by $DTW(F_D(60), F_j(60))$ where,

$$DTW(F_D(t), F_j(t)) = D[F_D(t), F_j(t)] + \min\{DTW[F_D(t-1), F_j(t-1)], DTW[F_D(t-1), F_j(t)], DTW[F_D(t), F_j(t-1)]\} \quad (17)$$

Thus if two multidimensional sequences, A and B, are identical, the DTW(A,B) is 0. The minimum cost path would be the diagonal of the distance matrix. A smaller value of the minimum cost path implies a greater similarity between the time series.

The number of profiles is given by the rule defined by (18). It determines the number of profiles by enforcing the probability of the least similar historical TAF is greater than the minimum probability threshold, $P_{min}$. A lower $P_{min}$ value will make $n$ large and a larger $P_{min}$ value would make $n$ small.

$$n = \underset{k}{\operatorname{argmax}} \frac{\frac{1}{DTW(F_D(60), F_{[j]}(60))^{DF}}}{\sum_{j=1}^{k} \frac{1}{DTW(F_D(60), F_{[j]}(60))^{DF}}} \geq P_{min} \quad (18)$$

Where,

$$DTW\left(F_D(60), F_{[1]}(60)\right) \leq DTW\left[F_D(60), F_{[2]}(60)\right) \leq .. \leq DTW(F_D(60), F_{[N]}(60)) \quad (19)$$

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

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

The profile probabilities are obtained after normalizing the minimum cost path for the '$n$' days.

$$P(k) = \frac{\frac{1}{DTW(F_D(60), F_{[k]}(60))^{DF}}}{\sum_{j=1}^{n} \frac{1}{DTW(F_D(60), F_{[j]}(60))^{DF}}} \quad (20)$$

We refer to capacity profiles obtained from this procedure as ***DTW Profile***.

### *D. A design-of- experiments approach to DTW Profile*

Constructing probabilistic capacity profiles from ***DTW Profile*** requires three input variables which determine the probabilities, the set of profiles and the degree of similarity. These three variables are Weighing Factor (WF), Dimension Factor (DF) and the minimum probability threshold ($P_{min}$). Since the generated set of probabilistic capacity profiles influence the ground delay decisions these parameters indirectly control the total realized cost. It is therefore essential to determine values of these variables which minimize average TC for an airport. We have determined input values which minimize the average TC for each airport using a technique called Response Surface Methodology used in design-of-experiments.

Response Surface Methodology (RSM) is a statistical technique which iteratively changes the inputs to determine an input combination which minimizes the output [15]. Inputs at every iteration change by moving in the direction which minimizes the output. Eventually, as the iterations increase, RSM determines the input combination or the region of input space which produces the minimum output. *In this research, we minimize the average TC, which is the output, by determining the values of the three input variables.*

Initializing the RSM, requires a factorial design and some starting values for the input variables where the output is evaluated. We have chosen a Face Centered Cube (FCC) factorial design which evaluates the average TC at 15 points (8 corners of the cube, 6 centers of the face and 1 centre of the cube) as shown in Fig. 1. The three dimensions of the cube represent the three variables. A FCC cube measures the output at three different levels for each input variable (a high value corresponding to 1, mid value corresponding to 0 and a low value corresponding to -1). Thus each of the 15 points represents a unique input combination to be evaluated. Depending on the input combination, the TC is evaluated for each day using the model in section III and averaged over all the days in the sample, to determine the average TC.

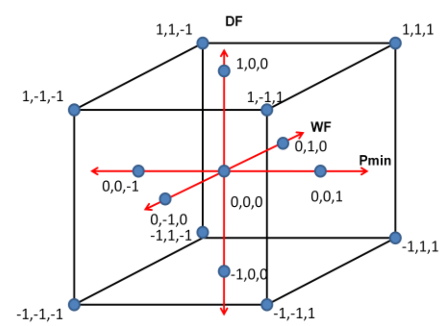



Figure 1. Face Centered Cube design

After determining the average total realized costs at the 15 points, the cube gets re-centered on the point which gives the lowest cost. This process continues till the minimum cost point converges at the center of the cube and the cube can't be re-centered furthered. Therefore this center point is the input combination that minimizes the average TC for that airport. We acknowledge that this approach is susceptible to a local minimum and we address this issue by randomly selecting multiple starting points and observe their convergence values.

### *E. STRATUS and Fog Clustering*

STRATUS is a program designed by MIT Lincoln Labs specifically for SFO to forecast the fog burn-off time and the probability that the fog would burn-off before 10am, 11am and noon. The forecast burn-off time is a proxy for the transition time from single landings to dual parallel landings at SFO. This burn off time is determined by using an ensemble of regression models and atmospheric boundary layer physics model. The probabilities are determined by comparing empirically the forecast time of burn-off with the actual time of burn-off at SFO [16]. STRATUS updates the forecast of the burn off time on an hourly basis from 2:00-11:00am PST. NASA Ames Research Center 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, actual burn off time and the probability that the fog would burn-off before 10am, 11am and 12 noon.

We base our analysis on the STRATUS forecast generated at 8:00am Pacific Standard Time (PST) for over 180 days in the summer months of 2004 to 2006. We choose the 8:00am forecast because it is the first forecast of the day incorporating the predictions from the Satellite Statistical Forecast Model (SSFM). We concentrated on the days when the fog burned off between 9:30am and 11:30am PST as the number of days outside this time bracket were very few. These days were binned in 15 minute periods according to the actual fog burn-off time between 9:30 to 11:30 a.m. PST. 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 III.

TABLE III. 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 |

From each bin we constructed a probabilistic capacity profiles 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). Each profile is a vector specifying the AAR value for each 15-minute time period from 7am to 10 pm. The profile $\boldsymbol{S_i}$ is determined by the average of AAR profiles in $\boldsymbol{B_i}$

$$\boldsymbol{S_i} = \left[ \sum_{h=1}^{d_i} \boldsymbol{A_h^{B_i}} \Big/ \boldsymbol{d_i} \right] \quad \boldsymbol{i \in 1..8} \tag{21}$$

The profiles are shown in the Fig. 2.

A closer inspection of the periods when the fog burns off reveals that there is a transition period lasting approximately for 45 minutes when the AAR is 45/hour. There is not an immediate increase in the AAR from 8 arrivals per period to 15 arrivals per period as assumed in [7]. 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.

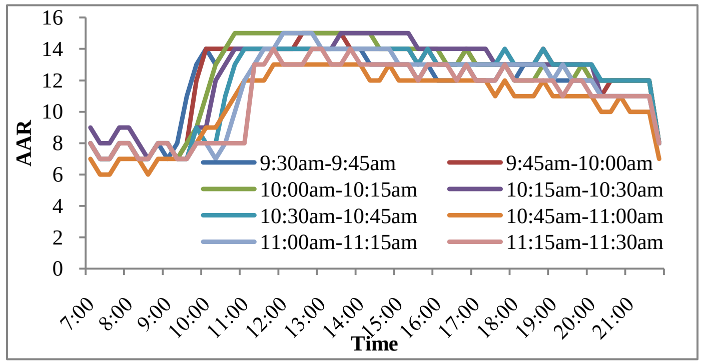



Figure 2. Profiles from Fog Burn off clustering

In Fig. 2, we observe an oscillation in the profiles. This is because of the way the AAR is reported in the ASPM database. The original rates are reported on a per-hour basis, which is then decomposed into 15-minute values in a manner that preserves integrality. Thus, the AAR of 60/hour reported as 15,15,15,15/quarter hour, an AAR of 45/hour as 10,11,12,12/quarter hour and an AAR of 30/hour as 8,7,8,7/quarter hour, causing the observed oscillation.

MIT Lincoln labs, on recommendation by the Traffic Management Unit at Oakland center, incorporated "risks" to the output of STRATUS. The risks are, in effect, cumulative distribution function (CDF) values of the form: ***P(Burn off*** $<$ $\boldsymbol{T_1}$***)=***$\boldsymbol{P_1}$, ***P(Burn off*** $<$$\boldsymbol{T_2}$***)=***$\boldsymbol{P_2}$ and ***P(Burn off*** $<$$\boldsymbol{T_3}$***)=***$\boldsymbol{P_3}$ where $\boldsymbol{T_1} < \boldsymbol{T_2} < \boldsymbol{T_3}$ and $\boldsymbol{P_1} \le \boldsymbol{P_2} \le \boldsymbol{P_3}$. The "risks" output by STRATUS for the day-of-operation can determine the probability of the eight profiles. Using the STRATUS-provided Cumulative Distributed Function (CDF) values, we linearly interpolate to obtain CDF values for each 15-minute period between 9:30 and 11:30 am.

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

$$\overline{\boldsymbol{P_{B_i}}} = \boldsymbol{Prob(Burn\,off} \le \lceil \boldsymbol{B_i} \rceil) - \boldsymbol{Prob(Burn\,off} \le \lfloor \boldsymbol{B_i} \rfloor) \quad \boldsymbol{i} \in [\boldsymbol{1..8}] \tag{22}$$

Where $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ are the lower and upper bin boundaries. Equation (22) 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 profile, $\boldsymbol{S_i}$, depicting burn off in $\boldsymbol{B_i}$, would have a probability

$$\boldsymbol{P_i} = \frac{\overline{\boldsymbol{P_{B_i}}}}{\sum_{i=1}^{8} \overline{\boldsymbol{P_{B_i}}}} \quad \forall \boldsymbol{i} \tag{23}$$

Equation (23) is a simple renormalization of the probabilities of the bins. The renormalized probabilities would sum to one.

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 burn-off time clustering***.

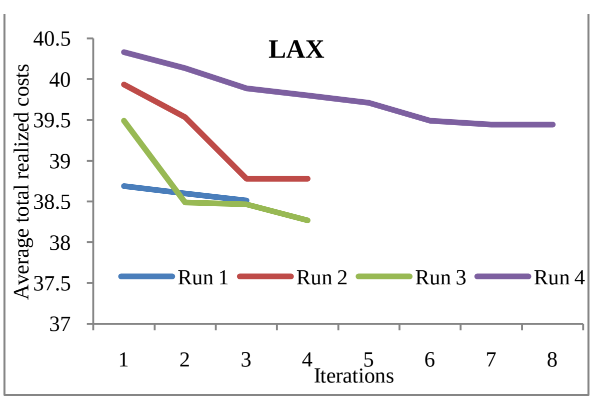
## V. CASE STUDIES AND COST COMPARISONS

The optimal number of cluster and probabilities for Naïve clustering and TAF based Clustering for all the three airports is shown in Table IV. The Naïve and TAF profiles are shown in Fig. 4 and Fig. 5 respectively.

TABLE IV. PROBABILITIES AND NUMBER OF PROFILES

| **Airport** | **# of Naïve profiles ($l^*$)** | **Probabilities of Naïve profiles ($P_i$)** | **# of TAF Clusters ($l^*$)** | **# of AAR profiles in $c_1$ ($k_1^*$)** | **Probability of profiles in $k_1^*$ ($P_{1,r}$)** | **# of AAR profiles in $c_2$ ($k_2^*$)** | **Probability of profiles in ($P_{2,r}$)** |
|---|---|---|---|---|---|---|---|
| **BOS** | 4 | 0.21, 0.36, 0.11, 0.32 | 2 | 2 | 0.45,0.55 | 2 | 0.61, 0.39 |
| **LAX** | 2 | 0.67, 0.33 | 2 | 2 | 0.36,0.64 | 2 | 0.29,0.71 |
| **SFO** | 3 | 0.25, 0.37, 0.38 | 2 | 2 | 0.46,0.54 | 3 | 0.22,0.41,0.37 |

We simulated ground delay strategies for 45 historical days from 2004 to 2006 for all the three airports using the GDP model described in section III. For SFO, we considered the days when the low lying marine stratus was observed. For these days at SFO, we generated the probabilistic capacity profiles for using both the 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. When applying the TAF based clustering method, we first determined in which of the TAF clusters a given day belonged to and then applied the profiles and probabilities under that particular TAF cluster in the GDP model. For example, a given day in SFO would either have two or three probabilistic capacity profiles depending on the classification of its TAF. For DTW Profiles using RSM, Fig 3 show the decrease in the average TC as the iterations increase for different starting points for the three airports. As expected the average total realized costs decrease with an increase in iterations.



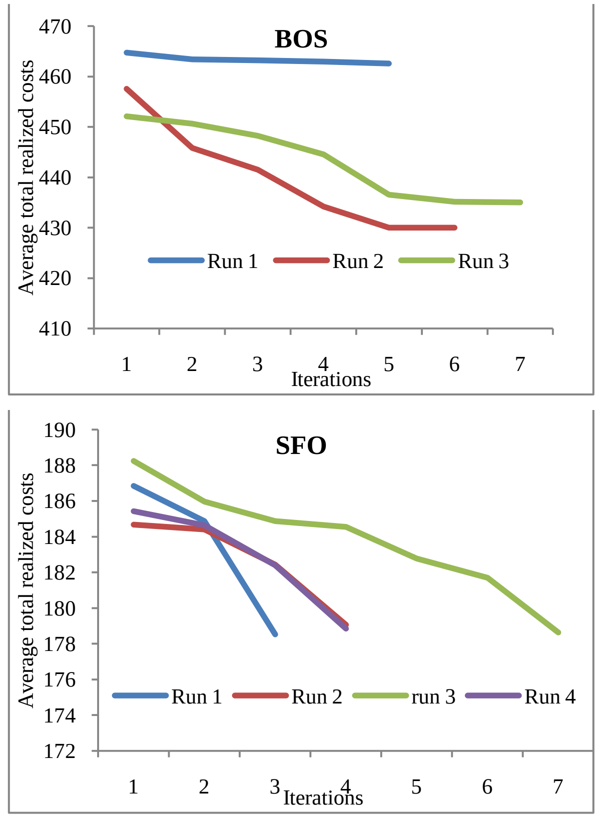




Figure 3. Decrease in average TC using RSM

The various total costs of delay are compared to a ***Perfect Information*** (PI) case where the air traffic managers 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 realized AAR profile with 100% probability of occurrence in the GDP model. With perfect information, we can eliminate all airborne holding while keeping ground holding to a minimum. The average total realized costs are given in the Table V with the standard deviation in the brackets.

TABLE V. TOTAL AVERAGE COST

| **Airport** | **PI** | **Naïve** | **TAF** | **DTW** | **Fog-Burn off** | **Naïve/DTW** |
|---|---|---|---|---|---|---|
| **LAX** | 20.41 (24.53) | 41.44 (58.73) | 41.75 (58.71) | 38.26 (60.29) | - | 1.10 |
| **BOS** | ***196.15 (335.41)*** | 616.64 (745.19) | ***570.02 (883.19)*** | ***429.97 (637.84)*** | - | 1.43 |
| **SFO** | ***96.5 (54.93)*** | 236.23 (156.7) | ***194.45 (145.28)*** | ***178.52 (102.74)*** | ***182.2 (109.73)*** | 1.32 |

We performed paired t-tests where the null hypothesis assumes the difference between the total costs obtained from the methodologies using the weather forecast and from naïve clustering is zero while the alternative hypothesis is the difference is other than zero. The values in ***bold italics*** indicate the cases where the null hypothesis is rejected i.e. the difference non- zero at a significance level of 0.1. The table also shows that using TAF to generate the profiles offers 10% to 40% cost reduction in simulated GDPs for the different airports.

The above table 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. For LAX, the average costs for the Naïve, TAF and DTW Profile methodologies are statistically equivalent, even though the cost from DTW Profiles is the lowest. For SFO, the DTW Profiles gives the minimum average cost of delays. This cost is marginally lower than the average delay from the STRATUS forecast. The two costs are statistically equivalent. The profiles derived from the STRATUS forecast yields the roughly the same level of costs TAF-based DTW method with RSM. This suggests, of the TAF-based methods, the DTW method in conjunction with RSM is the most promising for application at other airports, where, of course, the STRATUS forecast is unavailable. It can be stated that inclusion of weather forecasts in decision making leads to lowering of average total delay costs.

## VI. CONCLUSIONS

In this paper we have demonstrated how to employ weather forecasts to generate day-of-operation probabilistic capacity profiles. This represents steps towards the incorporation of weather forecast information to support probabilistic decision making in NEXTGEN, which can be taken without the expensive development of specialized weather forecast products. The TAF-based methodologies can be applied to any airport. It is shown that incorporating day-of-operation weather forecast information to plan the day-of-operation arrivals leads to a reduced realized cost when compared to profiles that do not make used of this information. 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 is focused. The careful use of the TAF offers a similar level benefit in GDP planning as a dedicated tool designed at considerable expense specifically for SFO.

## REFERENCES


[1] Mukherjee, A. & Hansen, M., 2007. A Dynamic Stochastic Model for the Single Airport Ground Holding Problem. Transportation Science, 41(4), pp.444-56

[2] Clark, D.A., 2009. Investigating a New Ground Delay Program Strategy for coping with SFO Stratus. In 89th AMS Annual Meeting ARAM Special Symposium on Weather. Phoenix, 2009.

[3] Ball, M.O., Hoffman, R., Odoni, A.R. & Rifkin, R., 2003. A stochastic integer program with dual network structure and its application to the ground-holding problem. Operations Research, 51(1), pp.167-71

[4] Richetta, O. & Odoni, A., 1993. Solving optimally the Static Ground Holding Policy Problem in air traffic control. Transportation Science, 27(3), p.228–238.

[5] Liu, P.B., Hansen, M. & Mukherjee, A., 2008. Scenario-based air traffic flow management: From theory to practice. Transportation Research Part B: Methodological, 42(7-8), pp.685--702.

[6] Kaut, M. & Wallace, S.W., 2003. Evaluation of Scenario-generation methods for Stochastic Programming. [Online] Humboldt-Universität zu Berlin, Humboldt University Berlin, Germany Available at: http://edoc.hu-berlin.de/series/speps/2003-14/PDF/14.pdf [Accessed 10 November 2010].

[7] Cook, L.S. & Wood, B., 2009. A Model for determining Ground Delay Program Parameters Using a Probabilistic Forecast of Stratus Clearing. In Eighth USA/Europe Air Traffic Management R&D Seminar. Napa, 2009.

[8] Klein, A., Jehlen, R. & Liang, D., 2007. Index with queuing component for National Airspace System performance assessment. In Seventh USA/Europe Air Traffic Management R&D Seminar. Barcelona, 2007.

[9] Buxi, G.S. &Hansen,M. Generation day-of-operation probabilistic capacity profiles from weather forecasts. In fourth International confrence on research in Air Transporttion, Budapest Hungary 2010.

[10] SAS Institute, 2004. SAS/STAT 9.1 User's Guide. Cary: SAS Institute.

[11] Kaufman, L. & Rousseeuw, P.J., 2005. Finding groups in data: an introduction to cluster analysis. WileyBlackwell.

[12] Jolliffe, I., 2005. Principal component analysis. Encyclopedia of Statistics in Behavioral Science.

[13] Ding, C. & He, X., 2004. K-means clustering via principal component analysis. In Proceedings of the twenty-first international conference on Machine learning. Banff, Alberta, 2004.

[14] Vlachos, M., Hadjieleftheriou, M., Gunopulos, D. & Keogh, E., 2003. Indexing multi-dimensional time-series with support for multiple distance measures. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. Washington DC, 2003.

[15] Montgomery, D.C. 2008. Design and analysis of experiments. John Wiley & Sons Inc.

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


## AUTHOR BIOGRAPHY

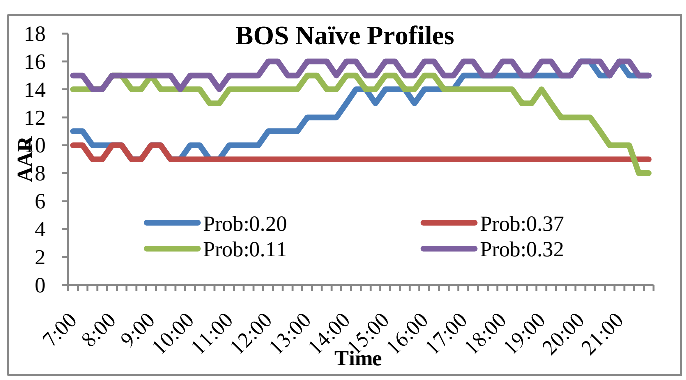
**Gurkaran Buxi** received his BSc. in civil and environmental engineering from Delhi College of Engineering, New Delhi, India in 2007 and his MSc in civil and environmental engineering from University of California, Berkeley in 2008.

He is a PhD candidate and a graduate student researcher in the civil engineering department at University of California, Berkeley. He is also a MA candidate in the department of statistics at the University of California, Berkeley. His research interests include ATFM, airport capacity and delays, optimization theory and machine learning.

**Mark Hansen** received his B.A. in physics and philosophy from Yale University in New Haven, Connecticut in 1980, and his Masters in city and regional planning (1984) and his Ph.D. (1988) in engineering science from the University of California, Berkeley.

He is a Professor of Civil and Environmental Engineering at the University of California, Berkeley. His research focuses on transportation planning, air transportation ATFM systems. He has taught graduate and undergraduate transportation courses in economics, systems analysis, planning, probability and statistics, and air transportation systems analysis, planning, probability and statistics, and air transportation.

Prior to graduate school, he worked as a physicist at the Environmental Protection Agency. He is the Berkeley co-director of the National Center of Excellence in Aviation Operations Research, a multi-university consortium sponsored by the Federal Aviation Administration.



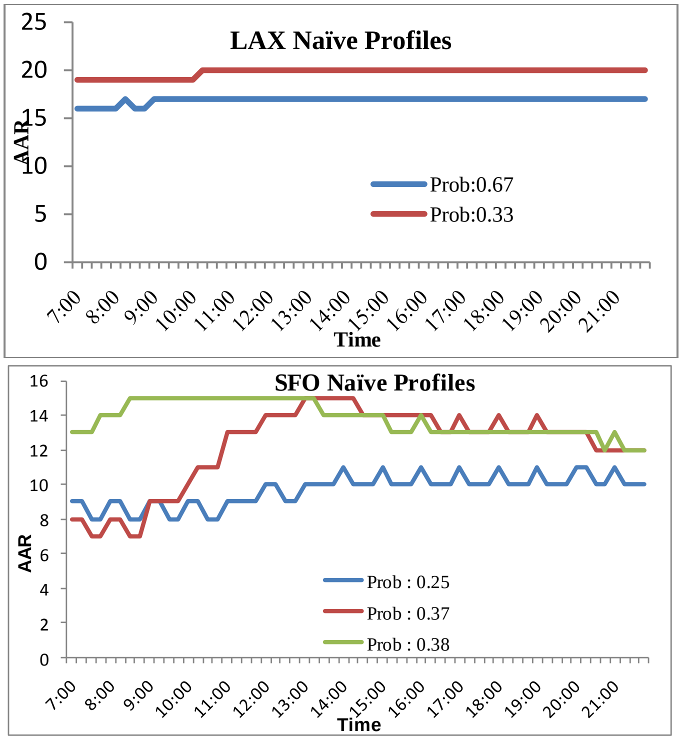
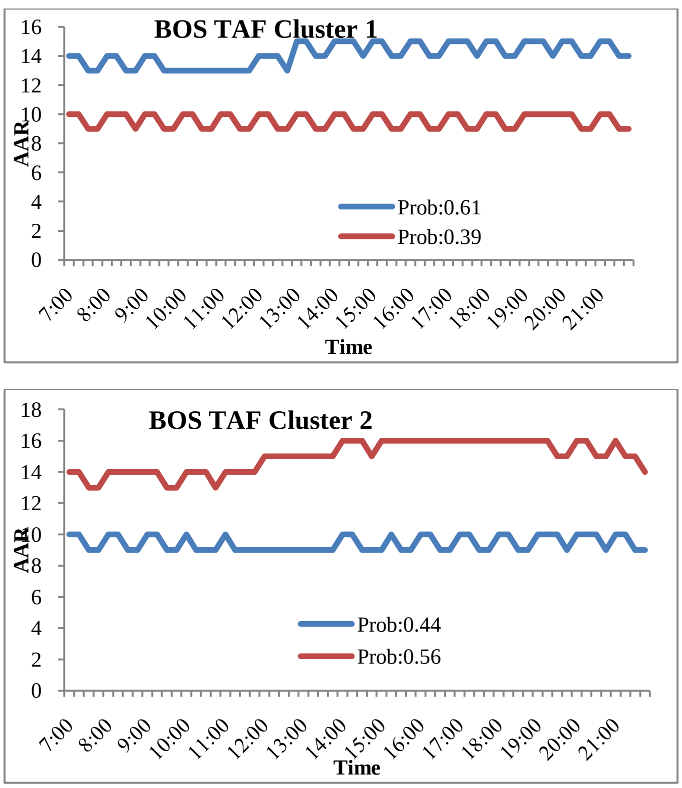



Figure 4. Naive Profiles for BOS,LAX and SFO



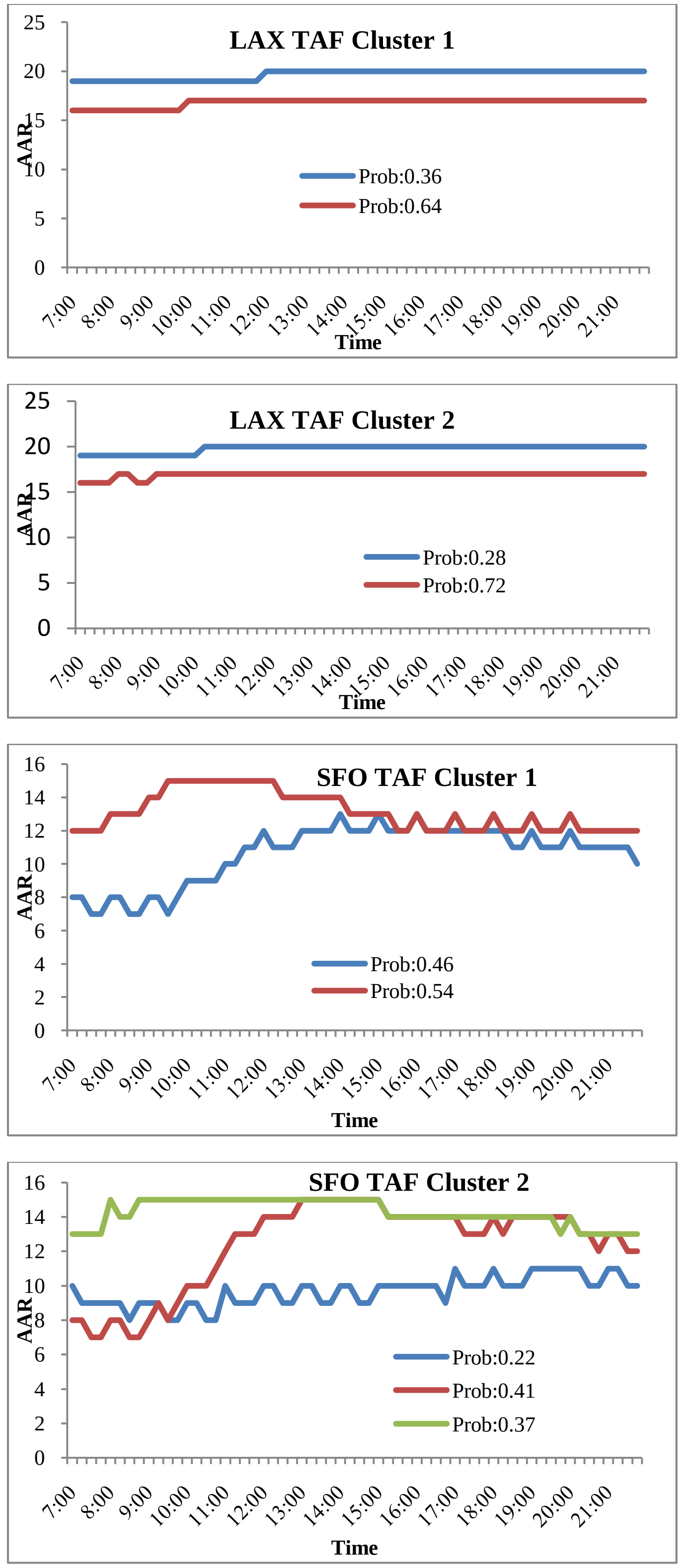




Figure 5. TAFClustering Profiles for BOS,LAX and SFO