**Title**
Generating day-of-operation probabilistic capacity scenarios from weather forecasts



**Author**
Buxi, Gurkaran

**Publication Date**
2012

Peer reviewed|Thesis/dissertation

# Generating day-of-operation probabilistic capacity scenarios from weather forecasts

By

Gurkaran Singh Buxi

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Civil and Environmental Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:
Professor Mark Hansen, Chair
Professor Samer Madanat
Professor Joan Walker
Professor Pravin Varaiya

Fall 2012

Abstract

Generating day-of-operation probabilistic capacity scenarios from weather forecasts

by

Gurkaran Singh Buxi

Doctor of Philosophy in Engineering – Civil and Environmental Engineering

University of California, Berkeley

Professor Mark Hansen, Chair

Airport arrival capacity, referred to here as the airport acceptance rate (AAR), is strongly influenced by the weather in the vicinity of the airport and thus AAR prediction necessitates an airport-specific weather forecast. Weather forecasts, however, are seldom accurate in predicting the actual weather conditions. Strategic decisions, for example arrival rates in a ground delay program (GDP), must be made ahead of time, usually more than two hours, when there is an uncertainty about the future capacity. This research uses probabilistic capacity scenarios to represent the uncertainty in the future arrival capacity. A probabilistic capacity scenario is defined as a time series of AAR values with which a certain probability of realization is associated. A set of probabilistic capacity scenarios may be used to represent the uncertainty in arrival capacity at an airport over the course of the day.

There has been considerable research in developing GDP models that determine efficient ground delay decisions and require probabilistic capacity scenarios as inputs. It is assumed that the capacity scenarios can be developed from weather forecasts or can be obtained from the expertise of the air traffic managers. There is, however, considerably less literature on the development of specific day-of-operation probabilistic capacity scenarios from weather forecasts. This limits the use of these GDP models in real- world application. This thesis fills that gap and presents methodologies to generate probabilistic capacity scenarios from weather forecasts.

In this thesis we develop methodologies for generating probabilistic capacity scenarios using a widely available airport-specific weather forecast called the Terminal Aerodrome Forecast (TAF). These methodologies require the issued TAF forecast and the realized capacity for days in the past. We apply and assess the performance of these methodologies on four US airports: San Francisco International Airport, Boston Logan International Airport, Chicago O'Hare International Airport and Los Angeles International Airports. Though we have focused on these airports as case studies, the TAF-based scenario generation techniques can be applied to any airport.

In the first methodology, TAF Clustering, the scenarios are representative capacity profiles for days having similar TAFs. Groups of similar TAFs are found using K-means clustering and the

number is verified using Silhouette value. In the second methodology, Dynamic Time Warping (DTW) Scenarios, the scenarios are the actual realized capacity profiles for days that have similar TAFs. The similarity between TAFs is determined using a statistical technique for comparing multidimensional time series called DTW.

DTW Scenarios uses three airport specific input parameters. These parameters control the numbers and the probabilities of the scenarios. We determine the values of the parameters through optimization to maximize the performance of the scenarios through minimizing average delay costs. The optimal values are determined through a specialized algorithm designed for situations where evaluating the objective function is computationally expensive.

For San Francisco International Airport we also use another forecast: the San Francisco Marine Initiative forecast (STRATUS) to develop the scenarios. In this methodology called, Fog burn-off clustering, the scenarios are representative capacity profiles for days that have the fog burn-off time in the same quarter hour.

We measure the efficacy of the various scenario generation methodologies in a real world setting based on 45 historic days for each of the four case-study airports. For each day, the generated scenarios are provided as inputs to a static stochastic ground delay model (SSGDM) that determines the series of planned arrival rates that minimize the sum of ground delay costs and expected air delay costs, assuming that the plan is not adjusted to evolving information. The ground delay is determined directly from the SSGDM whereas the realized air delay is determined from a queuing diagram based on the planned arrival rate and the realized arrival capacity. The realized delay costs are averaged over 45 days for each airport, and is the metric used to compare the different scenario generation methodologies. Employing this approach, we compare the different methods for capacity scenario generation against each other and against two other reference cases. Under the first reference case, Naïve Clustering, the scenarios are developed from historical capacity data without the use of the weather forecast. Groups of similar arrival profiles are determined though K-means clustering. In the second reference case, Perfect Information, we assume that the GDP is planned based on perfect information about the future arrival capacity.

Our results show that, on average, scenarios generated using the TAF-based DTW method results in the lowest delay cost amongst all scenario based methodologies. It is shown that capacity scenarios generated using day-of-operation weather forecasts can reduce the cost of delays by 5%-30% compared to scenarios that do not make use of weather forecast. The benefit of the TAF based approach is more pronounced on days that have a greater capacity-demand imbalance when compared to Naïve Clustering.

This thesis is dedicated to my Darji, Lochan Singh Buxi.

## Table of Contents

List of Tables ..... iv
List of Figures ..... v
Acknowledgements ..... vii
Chapter 1 Impact of terminal weather on airport capacity ..... 1
1.1 Introduction ..... 1
Chapter 2 Literature review ..... 5
2.1 Introduction ..... 5
2.2 Effect of weather on aviation ..... 5
2.2.1 Thunderstorms and convective weather ..... 5
2.2.2 Low visibility and low ceilings ..... 6
2.2.3 Wind speed and direction ..... 9
2.2.4 Precipitation and icing ..... 12
2.3 Air Traffic Flow Management (ATFM) models ..... 13
2.3.1 Determining the arrival capacity from historical data ..... 18
2.4 Static Stochastic Ground Delay Model ..... 18
2.5 Contributions to the literature ..... 20
Chapter 3 Scenario generation from weather forecasts ..... 24
3.1 Introduction ..... 24
3.2 Data for generating probabilistic capacity scenarios ..... 24
3.2.1 Airport arrival capacity data ..... 24
3.2.2 Terminal Aerodrome Forecast ..... 25
3.2.3 San Francisco Marine Stratus Initiative (STRATUS) ..... 27
3.3 Scenario generation using weather forecasts ..... 29
3.3.1 Notation ..... 29
3.3.2 TAF Clustering ..... 29
3.3.3 Fog Burn-off Time Clustering ..... 34
3.4 Reference cases ..... 39
3.4.1 Naïve Clustering ..... 39

3.4.2 Perfect Information (PI) ..... 40
3.5 Realized costs of delays ..... 40
Chapter 4 A design-of-experiment approach for scenario generation ..... 43
4.1 Introduction ..... 43
4.2 Dynamic Time Warping scenarios ..... 43
4.3 A design-of-experiment approach to determine parameter values ..... 48
Chapter 5 Results, conclusions and future work ..... 55
5.1 Introduction ..... 55
5.2 Probabilistic capacity scenarios from TAF clustering ..... 55
5.3 Probabilistic capacity scenarios from Naïve Clustering ..... 61
5.4 Optimal designs from dynamic time warping scenarios ..... 65
5.5 Average realized total cost of delay ..... 66
5.6 Conclusions, limitation and future research ..... 72
References ..... 75

## List of Tables

Table 2.1 Classifications of ATFM models ..... 14
Table 2.2 Dimensions of single destination airport, SATFM models ..... 15
Table 2.3 Overview of the scenario generation methodologies ..... 22
Table 3.1 Sample STRATUS forecast ..... 29
Table 3.2 Clusters, number of days and time boundaries according to fog burn-off time interval ..... 35
Table 3.3 Probability for burn-off in a quarter-hour period for 6-17-2004 ..... 38
Table 3.4 Probability for scenarios for 6/17/2004 ..... 39
Table 4.1 Dynamic time warping for two artificially generated data, WF=1 ..... 44
Table 4.2 Dynamic time warping for two artificially generated data, WF=10 ..... 45
Table 4.3 Summary of the three parameters for scenario generation ..... 48
Table 4.4 Upper and lower bounds for parameter values ..... 54
Table 5.1 Results of TAF Clustering for SFO ..... 55
Table 5.2 Results of TAF Clustering for LAX ..... 56
Table 5.3 Results of TAF Clustering for BOS ..... 56
Table 5.4 Results of TAF Clustering for ORD ..... 56
Table 5.5 Results of Naive Clustering for SFO ..... 62
Table 5.6 Results of Naive Clustering for LAX ..... 62
Table 5.7 Results of Naive Clustering for BOS ..... 62
Table 5.8 Results of Naive Clustering for ORD ..... 62
Table 5.9 Optimal design and number of scenarios for different airports ..... 65
Table 5.10 Average realized total cost of delay (in ground delay minutes) over 45 days ..... 67
Table 5.11 p-values from the paired t-test for SFO ..... 67
Table 5.12 p-values from the paired t-test for LAX ..... 68
Table 5.13 p-values from the paired t-test for BOS ..... 68
Table 5.14 p-values from the paired t-test for ORD ..... 68

## List of Figures

Figure 1.1 Revenue passenger enplanements versus Years ..... 2
Figure 1.2 Revenue departures versus Years ..... 2
Figure 1.3 Causes of National Aviation System Delays (Jan 2007-Dec 2007) Source: BTS ..... 3
Figure 2.1 National Transportation Safety Board (NTSB) related accidents by weather conditions (1994-2003) ..... 7
Figure 2.2 Ceiling and Visibility conditions for IMC and VMC at SFO ..... 9
Figure 2.3 Capacity distribution under IFR and VFR for SFO ..... 9
Figure 2.4 Wind speed and direction for 28L, 28R|1L, 1R for SFO ..... 10
Figure 2.5 Wind speed and direction for 28L, 28R|28L, 28R for SFO ..... 11
Figure 2.6 Wind speed and direction for 19L, 19R|10L, 10R for SFO ..... 11
Figure 2.7 Frequency of arrival capacity for different runway configurations ..... 12
Figure 2.8 Scenarios and realized capacity ..... 17
Figure 3.1 Arrival capacity profile hourly and quarter-hourly for SFO on 07/03/2007 ..... 25
Figure 3.2 Predicted versus actual burn-off time from STRATUS for SFO ..... 28
Figure 3.3 Distribution of days according to actual burn-off time ..... 28
Figure 3.4 An illustration of silhouette value calculation ..... 32
Figure 3.5 Silhouette value plots for 2 clusters and 3 clusters ..... 33
Figure 3.6 Capacity scenarios from Fog Burn-off time Clustering for SFO ..... 36
Figure 3.7 Capacity scenarios from Fog Burn-off time clustering for SFO for period around fog burn-off ..... 37
Figure 3.8 Linearly interpolated CDF for 6-17-2004 ..... 38
Figure 3.9 Realized capacity, PAAR and scheduled arrivals ..... 42
Figure 5.1 Probabilistic capacity scenarios for the days classified in $\boldsymbol{C_1}$ for SFO ..... 57
Figure 5.2 Probabilistic capacity scenarios for the days classified in $\boldsymbol{C_2}$ for SFO ..... 57
Figure 5.3 Probabilistic capacity scenarios for the days classified in $\boldsymbol{C_1}$ for LAX ..... 58
Figure 5.4 Probabilistic capacity scenarios for the days classified in $\boldsymbol{C_2}$ for LAX ..... 58
Figure 5.5 Probabilistic capacity scenarios for the days classified in $\boldsymbol{C_1}$ for BOS ..... 59
Figure 5.6 Probabilistic capacity scenarios for the days classified in $\boldsymbol{C_2}$ for BOS ..... 59
Figure 5.7 Probabilistic capacity scenarios for the days classified in $\boldsymbol{C_1}$ for ORD ..... 60
Figure 5.8 Probabilistic capacity scenarios for the days classified in $\boldsymbol{C_2}$ for ORD ..... 60
Figure 5.9 Probabilistic capacity scenarios for the days classified in $\boldsymbol{C_3}$ for ORD ..... 61
Figure 5.10 Scenarios from Naïve Clustering for SFO ..... 63
Figure 5.11 Scenarios from Naïve Clustering for LAX ..... 63
Figure 5.12 Scenarios from Naïve Clustering for BOS ..... 64
Figure 5.13 Scenarios from Naïve Clustering for ORD ..... 65
Figure 5.14 $\boldsymbol{TCD_{DTW} - TCD_{Naive}}$ versus $\boldsymbol{TCD_{PI}}$ for SFO ..... 70
Figure 5.15 $\boldsymbol{TCD_{TDTW} - TCD_{Naive}}$ versus $\boldsymbol{TCD_{PI}}$ for LAX ..... 71
Figure 5.16 $\boldsymbol{TCDT_{DTW} - TCD_{Naive}}$ versus $\boldsymbol{TCD_{PI}}$ for BOS ..... 71

Figure 5.17 $\boldsymbol{TCD_{TDTW} - TCD_{Naive}}$ versus $\boldsymbol{TCD_{PI}}$ for ORD..... 72

## Acknowledgements

I would like to thank my advisor, Professor Mark Hansen, for encouraging, pushing and guiding me in my doctoral pursuit. Over the course of my time in NEXTOR, I have been influenced by two characteristics of Prof. Hansen. Firstly, he will move hell and earth for his students. He might forget that he has to or there might be some delay but whenever I have been in an uncomfortable situation he was always there to "bat" for me. Secondly, his attitude towards research made me work efficiently and diligently. He has constantly encouraged me to write better.

I would like to thank Professor Samer Madanat, who enforced his "one question policy" in his class. This refined my thinking process to ask the most insightful question or as he said "the most bang for the buck". Professor Joan Walker's attitude made it effortless for me to approach her and discuss my research, academic and future career plans. Professor Pravin Varaiya's research suggestion made me think about the overall impact of my research in aviation. Some of the sections in this thesis are a direct consequence of our interactions.

I would not be in UC Berkeley had it not been for the interactions I had with Professor Geetam Tiwari in IIT Delhi. Professor Carlos Daganzo's and Professor Michael Cassidy's visit to India cemented by decision to come to Berkeley for my graduate studies.

My stay at Berkeley was enriched by great lab mates and colleagues. Bo Zou and Tasos Nikoleris were always there for me in NEXTOR through all my good and bad times. Amy Kim guided me on how to make killer presentations. Yoonjin Yoon and Jing Xiong always provided me with guidance on the art of doing research. Megan Smirti Ryerson, Gautam Gupta and Avijit Mukherjee provided informal guidance on my research area. I enjoyed playing Frisbee and talking with Elaine Kwan, Irene Kwan, Michael Seelhorst, Lu Hao, Matthew Elke and Yi Liu during lunch and coffee breaks. It was always fun interacting with Frank Ketcham and Giacomo Palombo both in NEXTOR and outside.

Berkeley provided me a set of great people with whom I shared my experiences. Transportation Engineering in McLaughlin is a close knit group. Josh Pilachowski, Offer Grembek, Vikash Gayah, Ilgin Guler, Eleni Christofa, Eric Gonzales, Celeste Chavis, Karthik Sivakumaran, Yiguang Xuan, DJ Gaker, Ipsita Banerjee, Aditya Medhury and Akshay Vij were the people with whom I had fun, worked, studied, learned and debated on a wide variety of topics.

Laura Marretta, Kyriakos Zarifis, Dilpreet Buxi, Mudit Narain, Madhur Mittal, Joseph Rajan, Kedar Hippalgaonkar, Ganesh Ananthanarayanan, Gopal Vaswani, Surya Prakash, Varun Boriah, Tagan Blake, Rohit Rai, Shing Tsoi, Rachit Kumar, Abhinav Golas, Shashwat Sehgal and Mohit Rajani were always there to ensure that I had fun outside McLaughlin.

Last but not the least, words would not be enough to thank my parent, Gurdeep Buxi and Tarvinder Bir Singh Buxi. Their unconditional love, support and affection gave me unparalleled confidence to pursue my own path without any compromise. My grandfather, Lochan Singh Buxi, made me the person I am today and I dedicate my thesis to him with all my love.

# Chapter 1 Impact of terminal weather on airport capacity

## 1.1 Introduction

Airline travel demand in the United States has significantly increased over the last 15 years. In 2010 the travel demand was roughly 30 percent higher than it was in 1996. Also, the past 15 years have witnessed two major shocks, the 9/11 attacks in 2001 and the financial recession in 2008, that temporarily abated the increase in demand. Fig 1.1 shows that the trend and the total passenger travel demand indicated by revenue passenger enplanements. The data shows that the demand recovered from the two shocks and is increasing again. The years 1996-2000, 2002-2007, and more recently 2009-2010 show an increase in the revenue passenger enplanements.

The airlines adjust their operations based on the travel demand and therefore, the number of operations in 2010 was 20% higher from 1996. Figure 1.2 shows the trend and the total number of aircraft revenue departures from 1996-2010. The total number of departures trend mimics the passenger demand trend for the same period. This indicates that the airline operations will probably increase in the future.

Though the passenger demand and airline operations have increased, the National Aviation System's (NAS) capacity has not kept pace. In 2007, the delays in the NAS reached their peak leading to substantial delays and cancellations but since then delays have temporary abated. The monetary impact of these delays on the economy, the airlines and to the passengers is estimated to be $32 billion in 2007 [**1**]. In 2007, delays due to weather constituted about 65% of the total delays and therefore are significant contributors to the total delay cost (Figure 1.3).

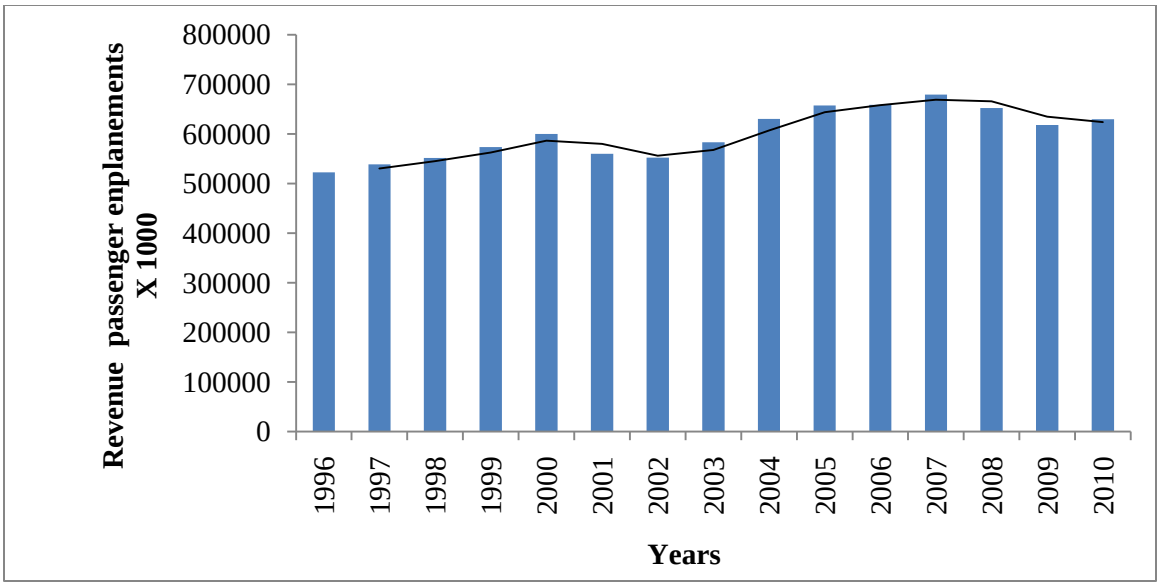


Figure 1.1 Revenue passenger enplanements versus Years

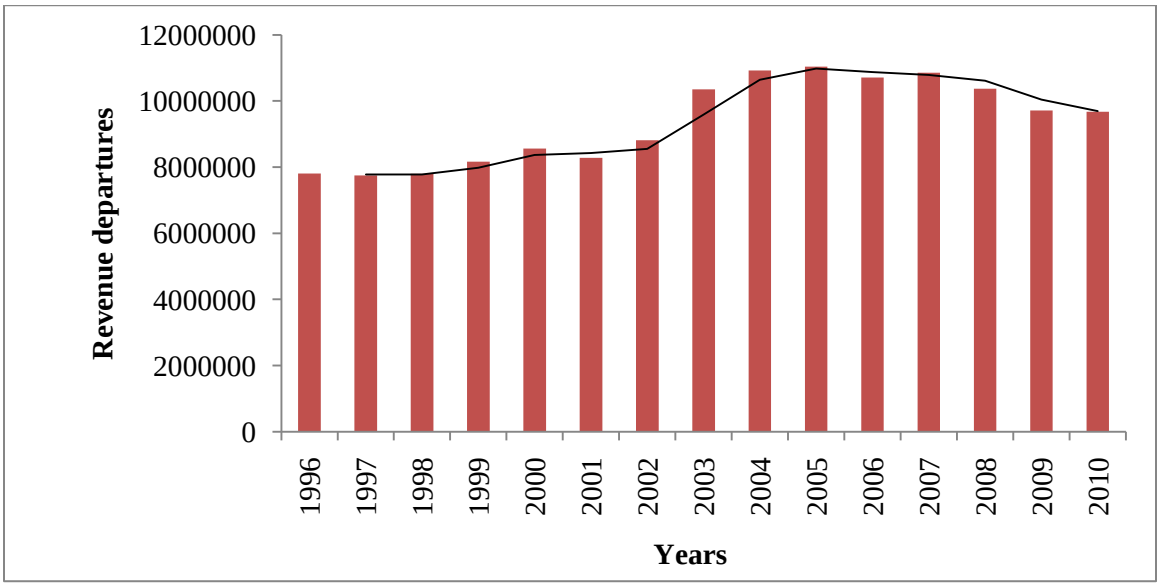


Figure 1.2 Revenue departures versus Years

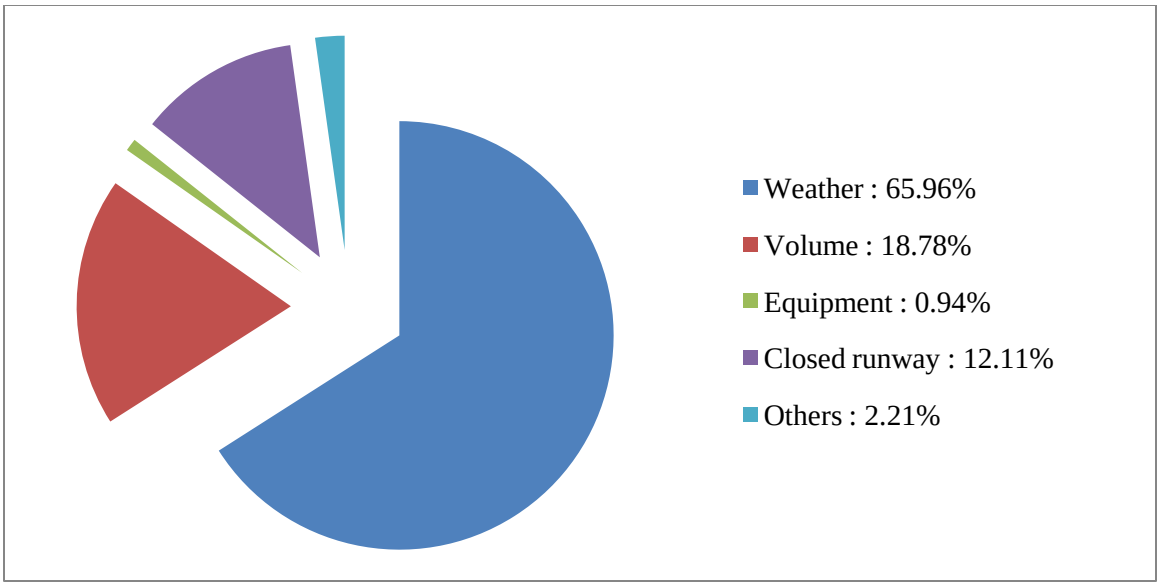


**Figure 1.3 Causes of National Aviation System Delays (Jan 2007-Dec 2007) Source: BTS**

The delays in the NAS can be alleviated by several potential methods while still serving existing and future demand. Firstly, new infrastructure that provides good weather capacity even in adverse weather can be introduced (for example by adding a runway to an existing terminal or increasing separation between parallel runways). This may be difficult or impossible to implement in space-constrained infrastructure like LaGuardia Airport (LGA). Moreover, modifications to the infrastructure may require high capital costs. Further the realized benefits of the capacity increase maybe different than the planned benefit, as the airlines and passengers would adjust their operations and travel behavior respectively. This is also referred as the feedback effects of induced demand. The benefit of expanding the aviation infrastructure has been investigated in greater detail in other research [**2**].

Secondly, the service provider (FAA) can mandate or encourage the operators (airlines) to upgrade the existing navigation systems and equipment to allow more precise and closely spaced aircraft operations. Precise GPS-based navigation, known as Required Navigation Performance (RNP), will enable aircraft to fly their preferred routes with high precision. Such an upgrade, combined with necessary procedures, would enable aircraft to fly with reduced separation standards thereby increasing the flow at metered fixes in the NAS (a fix could be merge point of several air lanes, a runway, etc.).

Finally, better weather forecasting can enable air traffic managers to more accurately predict airport and airspace capacities and make better decisions about when, where, and how to meter traffic in order to make best use of available capacity. One aspect of this is to explicitly recognize and characterize weather-related uncertainty in future capacity and employ probabilistic decision methods that optimize the expected performance of the system.

This research focuses on the last potential solution, to improve the strategic decision making of the service provider's by including the uncertainty in future weather conditions.

Weather conditions influence the capacity of NAS resources, such as airports and sectors. Strategic decisions rely on weather forecasts to predict the future weather conditions, and therefore the capacity of NAS resources, for aligning the traffic with predicted capacity to reduce the cost of delays in the NAS. Uncertainty about future weather and its capacity impacts results in uncertainty about the future capacity of the NAS resources. In the present system, this uncertainty is managed through air traffic manager judgment as well as input, provided in regular teleconferences, from flight operators. Air traffic managers make decisions based on their experience, weather forecasts, current weather, and projections of future demand. The current strategic decisions tend to be conservative and therefore increase costs in the NAS. There is no formal mechanism which converts the weather forecast and its associated uncertainty into metered rates.

In this research we will focus on managing capacity uncertainty in the airport terminal environment. In particular, we will develop methods for using weather forecasts to determine an optimal arrival rate in a manner that takes into account a terminal weather forecast for the day-of-operation. We characterize the evolution of capacity over the day and its uncertainty based on the forecast. We propose several methodologies for doing this. All of them capture uncertainty in the arrival capacity, referred here as the airport acceptance rate (AAR), through probabilistic capacity scenarios. A probabilistic capacity scenario is defined as a time series of AAR values with which a certain probability of realization is associated. A set of probabilistic capacity scenarios may be used to represent the uncertainty in arrival capacity at an airport over the course of the day. The scenarios are generated for four US airports using the weather forecast and the realized capacity profiles from May-September for the years 2004-2006. We develop a platform to assess the performance of the scenarios in a real world setting based on several days in the past. For each of these days, we will determine the realized cost of delays that is the sum of ground delay cost and realized air delay cost. For each airport, our metric to compare the different scenario generation methodologies the average of the realized delay costs over several days.

The rest of the thesis proceeds as follows. In Chapter 2 we provide the literature review. We discuss the influence of weather on capacity and the different Air Traffic Flow Management (ATFM) models. These models determine optimal strategic decisions to minimize the expected cost of delay. In Chapter 3, we discuss the terminal weather forecasts used in this research: San Francisco Marine Stratus forecast (STRATUS) and the Terminal Aerodrome Forecast (TAF). STRATUS is available only for SFO while the TAF is a forecast issued for all major airports. We present our methodology, called fog burn-off time clustering, to determine scenarios for SFO using the STRATUS. We then present our first methodology using the TAF called TAF Clustering for generating scenarios. We also present two reference cases: Naïve Clustering and Perfect Information to quantify the benefit of using weather forecasts. We also present our platform for comparing different scenario generation techniques. In Chapter 4, we introduce another methodology requiring the TAF called Dynamic Time Warping (DTW) Scenarios. Three parameters are introduced which enhance the performance of DTW Scenarios. We determine the optimal values of these parameters though an optimization framework. In Chapter 5 we present the results, offer conclusions and directions for future research.

# Chapter 2 Literature review

## 2.1 Introduction

This chapter presents a literature review pertinent to this thesis. The literature review is divided into two parts. The first part outlines the effect of weather on capacity and the second part discusses the Air Traffic Flow Management (ATFM) models, the uncertainty in capacity inputs and the Static Stochastic Ground Delay Model (SSGDM).

In the first part we show qualitatively the impact of weather on the NAS. We begin the second part by broadly classifying ATFM models if they account for uncertainty in the future state of the NAS resources and also on the number of airports included in the model. This is followed by a literature review on the modeling and prediction of arrival capacity, a required input in a number of the ATFM models. This research extensively utilizes the Static Stochastic Ground Delay Model and a section is devoted to discussing that model in greater detail. This chapter concludes with a discussion on the contribution of this research to the literature.

## 2.2 Effect of weather on aviation

Adverse weather affects the capacity, safety and efficiency of the National Aviation System (NAS). This section qualitatively explains the impact of weather on the NAS.

### 2.2.1 Thunderstorms and convective weather

Thunderstorms and convective weather (T&CW) cause almost half of the delays in the summer months in the United States [**3**], [**4**]. The high T&CW intensity regions pose several safety hazards to the aircraft in the vicinity and also decrease the efficiency in the NAS.

The hazards include high turbulence, intense up and downdrafts, lightening, hail and wind shear [**3**]. Damage to the aircraft from lightening and hail may render the aircraft unfit for operations and may increase the maintenance cost. Since the intensity and the location of the T&CW change with time, they make regions of the airspace, called sectors, having high storm intensity unflyable. This decreases the available flyable airspace. Aircraft scheduled to fly through these high intensity T&CW activity sectors have to deviate from their nominal routes which increases their travel time. The aircraft also have to carry additional fuel to account for the increase in the route. The increase in cost, increased passenger time and additional travel time decrease the efficiency of the NAS.

In times of severe T&CW, when the flyable airspace decreases it results in a capacity-demand imbalance. Aircraft originally scheduled to fly through the sectors having T&CW activity are diverted to fly through the unaffected, weather free sectors. These diversions increase the

demand of traffic through the weather free sectors as aircraft compete for the limited flyable airspace.

The capacity of a weather free sector is predominantly determined by the work load of the controller. The controller is responsible for managing and ensuring safe passage of aircraft though the sector. When more aircraft are diverted to fly through a sector it increases the work load of the controller. The diversion of aircraft and the increased work load of the controller constitute a capacity-demand imbalance. To mitigate the effects of this capacity-demand imbalance, the FAA initiates the Airspace Flow Program (AFP). The AFP is a type of an Air Traffic Flow Management (ATFM) program and it aligns the projected demand with the available capacity through the sector. Through the AFP, the FAA assigns "slots" or times when a particular aircraft should enter the capacitated sector. The rate at which the aircraft enter the capacitated sector is metered to ensure a smooth transition though the sector airspace. A particular aircraft might be delayed either at its origin airport or at some upstream holding pattern to ensure that it arrives at the sector at it assigned slot time.

### 2.2.2 Low visibility and low ceilings

Ceiling and visibility (C&V) are a major concern for air traffic decision makers since C&V play an essential role in controlling takeoffs, approaches, and landings. Low C&V conditions produce negative impacts on aviation and contribute to over 20% of all weather-related accidents in the U.S. civil aviation sector as shown in Figure 2.1 [**5**]. Ceiling and visibility directly affect the safety and efficiency of terminal operations. C&V conditions in the terminal area influence decisions on the opening and closing of individual runways [**6**]. The same value of ceiling or visibility may affect different airports differently for example, a particular value of ceiling and visibility at St. Louis Lambert Field may reduce the arrival rates, while the same value at Dallas-Fort Worth or Boston's Logan Airport may have no impact at all [**7**]. The effect of C&V conditions on an airport's runway is primarily determined by the runway geometry and on the landing instruments at the runway.

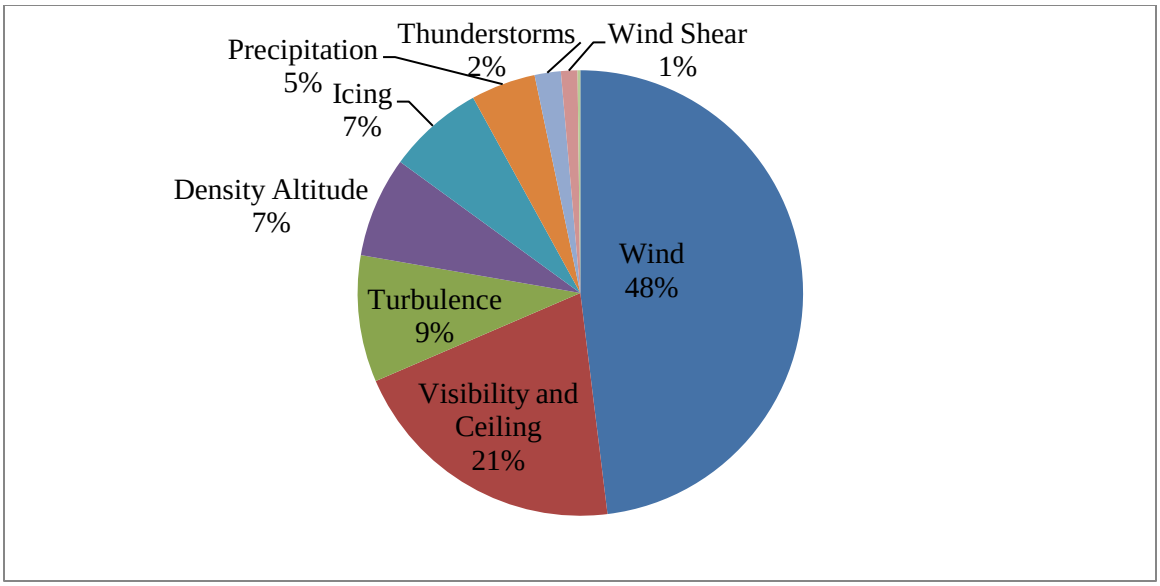


**Figure 2.1** National Transportation Safety Board (NTSB) related accidents by weather conditions (1994-2003)

Though the most critical impact of low C&V conditions is on the general aviation (GA) they also affect the commercial air traffic. Poor C&V conditions accounted for 24% of all GA accidents between 1989 and 1997. It is reported that in poor C&V conditions, GA pilots often lost control of the aircraft or collide with the terrain (controlled flight into terrain) as they lacked the proper certification or the aircraft did not have the necessary equipment required in such conditions [**8**].

C&V conditions, the topography of the airport and instrumentation of the runway affect the runway configuration and therefore influence the runway capacity. Poor C&V conditions reduce the arrival and departure rates for major US airports causing an increase in delays which affects the commercial air traffic.

The C&V conditions at the airport influence the weather category the airport operates in for the corresponding time period. The operating weather categories for any airport can broadly be classified as Instrument Flight Rules (IFR) and Visual Flight Rules (VFR). The IFR category typically corresponds to a low C&V values while the VFR category corresponds to good weather conditions implying a higher C&V values. When an airport operates in IFR category, the arriving aircraft to that airport are required to operate in instrument metrological conditions (IMC). The pilots navigate the aircraft using the instruments aboard the aircraft while maintaining separation as instructed from the air traffic controllers. In the VFR category, pilots can either opt for visual approaches (Visual metrological condition, VMC) i.e. navigate the aircraft by maintaining visual separation from the aircraft proceeding them or request to operate in IMC. When a trailing aircrafts pilot opts for visual approaches, the trailing aircraft's pilot takes responsibility to maintain adequate separation to mitigate the effects of the wake vortex from the proceeding aircraft. When pilots opt for visual approaches, it translates into higher throughput at the runway [**9**].

Furthermore when pilots opt for VMC approaches, FAA allows for parallel landings on pairs of parallel runways where the runway centerlines are separated by 700 feet. This is best illustrated in the case of San Francisco International Airport (SFO) which has two pairs of parallel runways

(28R-28L and 19R-19L) separated by 750feet. Under VMC approaches, the arrival rate is 15 arrivals per quarter hour and both the runways (28R-28L) are available. On the contrary, in poor C&V conditions, implying IMC conditions, only one runway (28R) is available for landing and the rate is reduced to 8 arrivals per quarter hour.

The phenomenon of the horizontal visibility reducing below 0.54 Nautical mile (≈1 KM) is referred as fog, in the terminal area. Thus, fog is defined as very low visibility in the terminal area. A ten year study of aviation accidents by the US National Transportation Safety Board (NTSB) indicated that 29% of all accidents in poor C&V conditions were caused by fog conditions [**5**]. Airports that have parallel runways, the occurrence of fog can reduce the arrival capacity by closing one of the parallel runways. Considerable research has been done to predict the time the fog would dissipate from the airport and therefore predicting the time the arrival capacity would increase.

When the arrival capacity an airport reduces due to poor C&V conditions, the FAA tries to align the arrival demand with the available capacity using some air traffic management initiatives. These initiatives could either be increasing miles in trail restriction between pairs of arriving aircraft or initializing a Ground Delay Program (GDP) at the affected airport. By increasing miles in trail restriction, the controllers increase the separation between the arriving aircraft and thereby reduce the arrival rate an airport. In a GDP, the flights bound for the affected airport are delayed on the ground at the origin airport. If it is known that an aircraft would incur air delay over the destination airport, then it is preferable to have that aircraft wait on ground at its origin to absorb the air delay. This translates air delays to the less expensive ground delays.

Figure 2.2 shows the C&V conditions and the corresponding weather operating category for SFO for November 2010, December 2010 and January 2011 [**10**]. It can be noted that SFO operates in the VFR category when the ceiling and visibility is greater than 2500 feet and 8 nautical miles respectively. While, anything less than the above values is classified as IFR. It should also be mentioned, for these dates SFO was operating in IFR category for 24 percent of time and in VFR category for 76 percent of time. Figure 2.3 shows the distribution of the arrival capacity or Airport Arrival Rate (AAR) for the two categories. When the airport operates in IFR category the arrival capacity is low whereas when the airport operates in the VFR category, the capacity is predominantly higher.

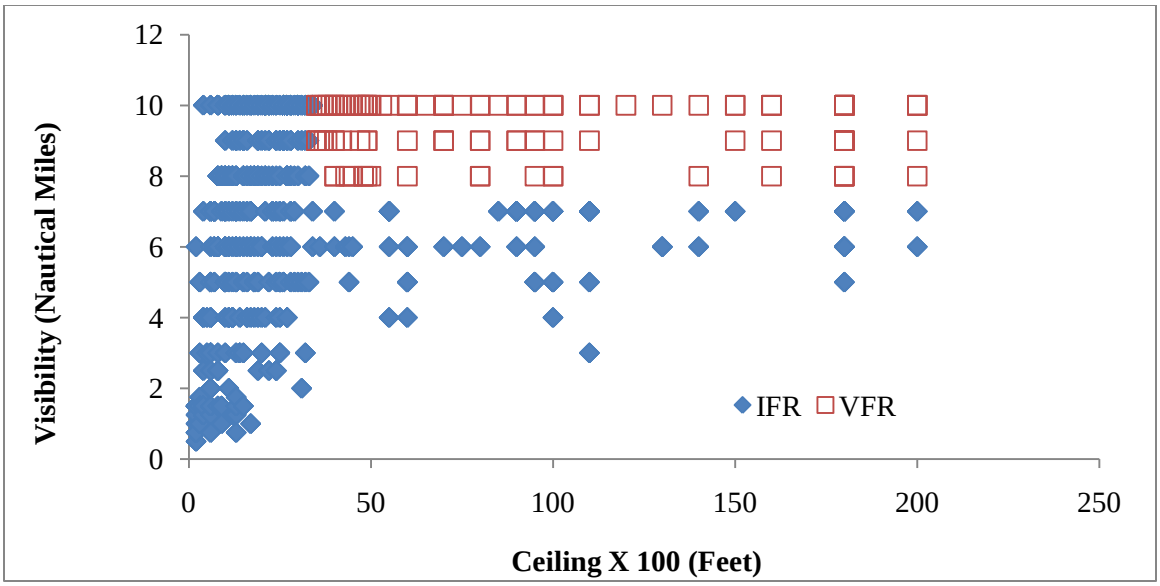


Figure 2.2 Ceiling and Visibility conditions for IMC and VMC at SFO

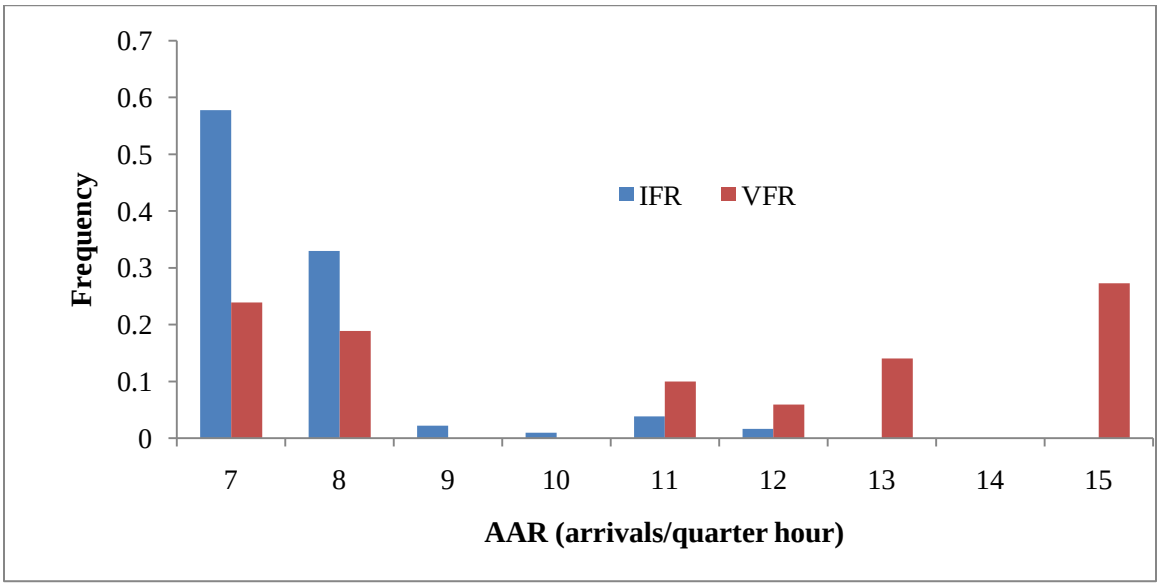


Figure 2.3 Capacity distribution under IFR and VFR for SFO

### 2.2.3 Wind speed and direction

Figure 2.1 show that winds caused a majority of accidents in 1994-2003. The prevailing winds in a particular period are an important factor as they determine the runway configuration and also determine the aircraft type that can use a particular runway. FAA has specified the maximum cross wind and tail wind velocities for different classes of aircraft using a runway. For example at San Francisco International Airport, under the most common weather conditions the arrivals

are preferred to land on the Runway 28 pair and departures are preferred on Runways 1 pair. If the wind direction is from the west and has a high velocity, it hinders departures on the Runway 1 pair and then the departures and arrivals both occur on Runway 28 pair (to adhere to the maximum crosswind limitation). When winds are from the south or east, as they often are during winter storms, the arrivals land on the Runway 19 pair and departures occur on the Runway 10 pair. If the winds from the east are too strong to allow arrivals on the Runway 19 pair, due to aircraft crosswind limitations, then both arrivals and departures have to use the Runway 10 pair. When winds from the north are too strong the aircraft have to land on the Runway 1 pair, and depart on either Runway 28 or 10 pair or at times the departures are mixed with the arrivals on the Runway 1 pair [**11**]. The wind speed and direction for the different runway configuration for November 2010, December 2010 and January 2011 for SFO are shown in Figure 2.4, 2.5 and 2.6. Each point on the chart shows the wind speed in East and North co-ordinates. The ray originating from any point to the center (0,0) shows the direction in which the wind was blowing.

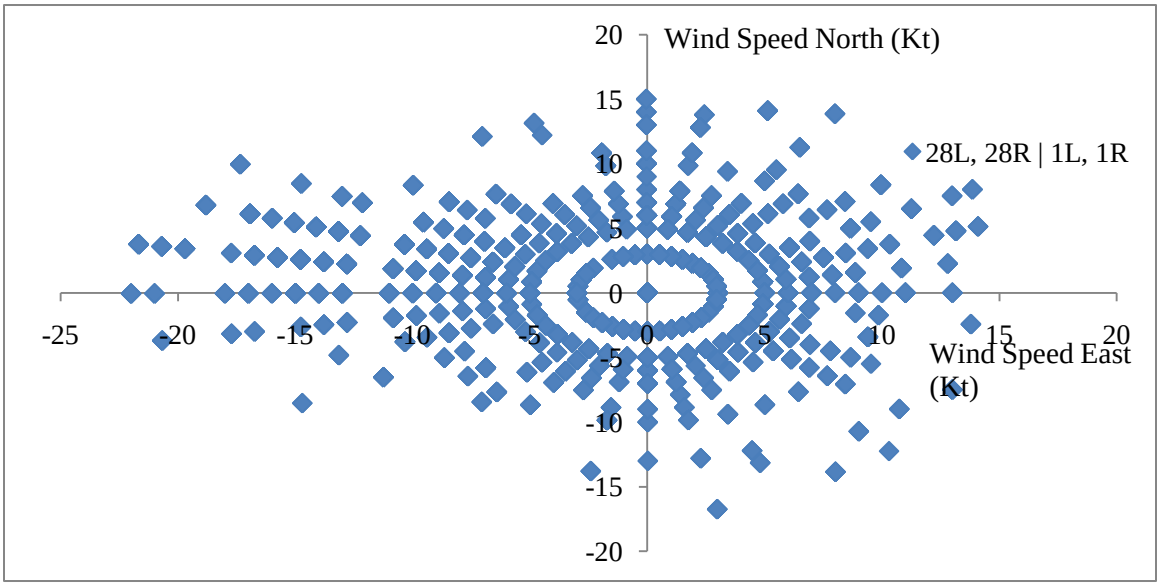


**Figure 2.4 Wind speed and direction for 28L, 28R|1L, 1R for SFO**

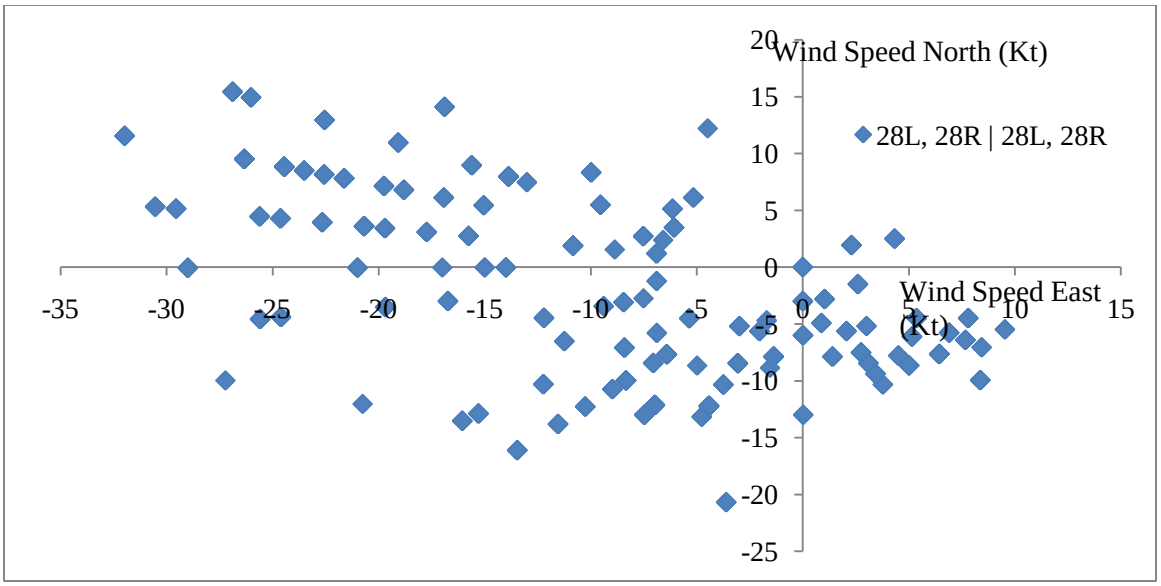


Figure 2.5 Wind speed and direction for 28L, 28R|28L, 28R for SFO

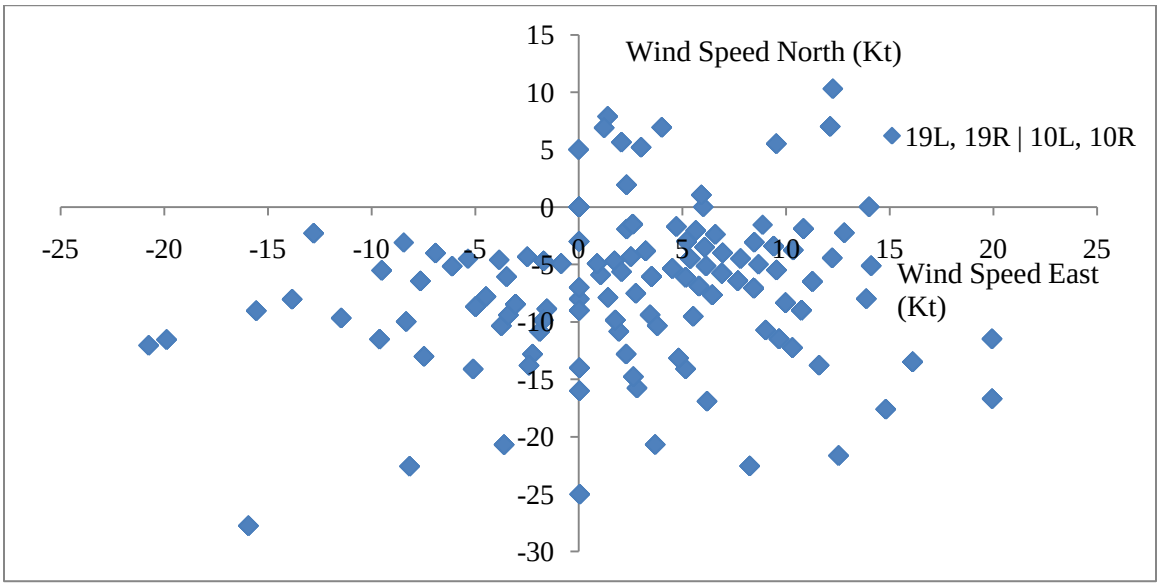


Figure 2.6 Wind speed and direction for 19L, 19R|10L, 10R for SFO

Figure 2.7 shows the airport arrival rate (AAR) per quarter hour for the different runway configurations. It can be seen that only the preferred configuration 28L, 28R|1L, 1R is able to accommodate a high arrival rate of 15 aircraft arrivals per quarter hour. While configurations 19L, 19R|10L, 10R and 28L, 28R| 28L, 28R accommodate the low and medium arrival rates respectively corresponding to 8 arrivals per quarter hour and 11 arrivals per quarter hour respectively.

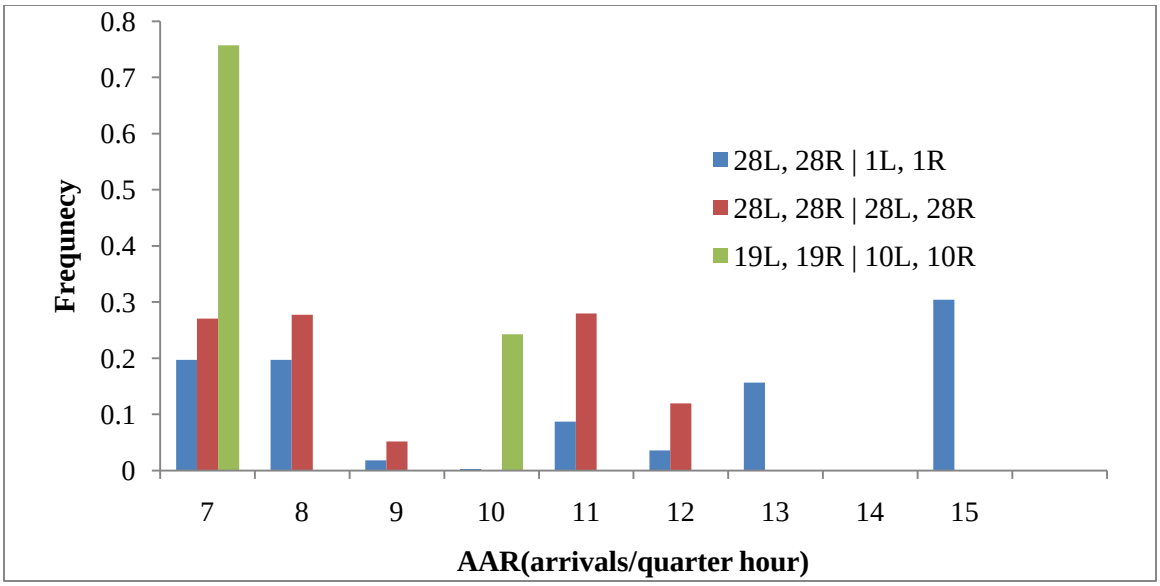


Figure 2.7 Frequency of arrival capacity for different runway configurations

### 2.2.4 Precipitation and icing

As shown by Figure 2.1 precipitation and icing caused over 11% of all accidents from 1994-2003. Precipitation and icing effect aviation in three ways: Firstly, they reduce the runway capacity as they are often accompanied with poor C&V. Secondly, they can damage the aircraft engine and air frame. Thirdly, they decrease the friction between the tires and ground for takeoff and landing.

Light precipitation generally has little or no affect on aviation operations but high precipitation is often associated with wind shear and turbulence. This may put stress on the airframe and the turbulence may cause discomfort to passengers.

Icing can be categorized as in-flight icing or ground icing. In-flight icing is dangerous to the aircraft and also reduces the efficiency of the NAS. Ice on the aircraft surface increases the weight of the aircraft and provides erroneous instrument reading. Ice on the engine though may cause engine failure and a loss of power. The 1994 crash of an American Eagle ATR-72, occurred because the aircraft encountered sudden icing conditions causing the pilot to lose control. Even a very small amount of ice on the wings increases the drag causing the aircraft to consume more fuel. Aircraft pilots avoid flying through regions which have icing conditions this increases their route length thereby increasing the onboard fuel requirement. This also delays the arrival of the aircraft at its destination and may cause a delay ripple effect decreasing the efficiency of the NAS.

Ground icing on the other hand, reduces the runway operations and the runway staff constantly has to work to prevent ice and snow buildup. In excessive snow and ice conditions the airport does not accept any arrivals and the aircraft attempting to land have to be diverted to other airports. This also causes delays and increases costs for the NAS users. Excessive snow and ice conditions may also render some gates and taxiways unusable. In December 2010, London

Heathrow Airport closed for almost two days due to excessive snow and ice conditions causing excessive flight delays, schedule disruptions and passenger discomfort.

The above discussion qualitatively shows that thunderstorm and convective weather, low ceiling and visibility, wind, precipitation and icing affect capacity. Thunderstorms and convective weather affect the en-route airspace by decreasing the sector capacity whereas the other weather conditions influence the terminal capacity. In this research we focus on the terminal and ignore the en-route, and therefore focus on the weather factors influencing terminal capacity.

## 2.3 Air Traffic Flow Management (ATFM) models

As mentioned in Section 2.2 adverse weather can have major impacts on aviation operations. When present in the vicinity of an airport, it decreases the operational capacity at that airport. Under such situations, if the scheduled demand exceeds the reduced capacity, air traffic flow specialists manage the air traffic demand by metering traffic into the affected airport. This metering of air traffic is achieved by air traffic flow management (ATFM) initiatives. ATFM initiatives align the air traffic demand with capacity, ensuring the available capacity is used efficiently and delays are absorbed in a safe, cost-effective manner. ATFM initiatives include ground holding, ground delays, rerouting aircraft to avoid areas of adverse weather, and increasing the miles-in-trail between pairs of aircraft.

ATFM initiatives can be divided into two categories: tactical and strategic. The tactical and strategic ATFM initiatives vary in the time between issuance and implementation. Tactical ATFM initiatives are implemented with a smaller lead time and therefore are issued when the future state of NAS resources (that includes the en-route sector capacity, travel times between sectors, airport capacity) can be predicted. These are implemented in order of seconds to minutes after they are issued. Examples of tactical ATFM initiatives include ground stops and miles-in-trail restrictions. During a ground stop, all the departing aircraft from an airport are ground-held for the duration of the stop and thus their departure is temporarily delayed. The miles-in-trail restrictions meter the arrival of aircraft at a specific en-route point, such as an entrance to a sector or en-route waypoints.

Conversely, strategic ATFM initiatives are issued several hours in advance when the future state of the NAS resources is difficult to predict. Strategic ATFM initiatives are required to ensure that the air traffic demand matches the future capacity, and to mitigate potential congestion and delays. When foreseen delays are unavoidable, strategic initiatives transfer the delays from one NAS resource to another, for example from the air to the airport surface, to reduce the associated delay costs. These initiatives work at a more aggregate level. The initiatives are usually issued between two to twelve hours before their implementation [**12**]. Examples of strategic ATFM initiatives include Ground Delay Programs (GDPs) and air space flow programs (AFPs).In a GDP, the aircraft bound for a particular destination airport that is experiencing a capacity-demand imbalance are delayed at the origin airports and given new, delayed departure times. Whereas, an AFP is issued when adverse weather constrains a region of the en-route airspace. Flights scheduled to fly through this airspace are given expected departure clearance times (EDCT) thereby metering the traffic through this airspace. Strategic initiatives are particularly

important for the airlines as they allow them to plan their operations further in advance (for example increase fuel intake in case of planned longer reroutes).

ATFM models found in the literature are optimization models seek to minimize total user cost while adhering to capacity constraints of NAS resources. Decision variables may include a combination of ground delays, link travel times, routings, and air holding. Depending on the ATFM model, it may require deterministic or probabilistic information about the future capacities of NAS resources as inputs. Deterministic ATFM models assume deterministic values for future capacities. These models thus assume that the future state of the NAS resources can be known with certainty. Deterministic ATFM models minimize the sum of ground delay and air delay costs. ATFM models that capture uncertainty in future capacities of NAS resources are known as stochastic ATFM models (SATFM). These models represent future capacities probabilistically. These models minimize the sum of expected ground delay costs and air delay costs.

ATFM models can also be divided based whether they consider individual airports or multi-airport networks. The single airport, ATFM models determine efficient ground delay decisions in a GDP for a single airport. The multi-airport ATFM models optimize ground delays for a network of airports, taking into account the propagation of network delay in downstream segments for connected flights. The single airport and multi-airport ATFM models may also consider the en-route portion of the flight. In this case, the models also require the capacity of en-route sectors and travel time between sectors. Decision variables in such models include airborne as well as ground delay, as well as routings. Table 2.1 classifies ATFM models based on whether they are deterministic or stochastic and whether they consider single airports or multiple airport networks. Single airport and multi-airport ATFM models that include the en-route information are denoted in bold italics.

**Table 2.1 Classifications of ATFM models**

| | **Single airport** | **Multi airport** |
|---|---|---|
| **Deterministic ATFM models** | [**12**] | [**13**]*;* [**14**]*;* [**15**]*;* [**16**]*;* [**17**] |
| **Stochastic ATFM models (SATFM)** | [**18**]**;** [**19**]**;** [**20**]**;** [**21**]**;** [**22**]**;** [**23**]**;** [**24**]**;** [**25**]**;** [**26**] | [**27**]*;* [**28**] |

SATFM models assume that the arrival capacity of the destination airport can be modeled as a random variable. These models can be static or dynamic, aggregate or disaggregate, and scenario-based or Markovian models. We explain the distinctions below. Table 2.2 classifies the more notable single airport SATFM models based on them.

Static SATFM models assume decisions are made at a single point in time and cannot be revised. Conversely, in dynamic models allow for decision making over time as more information becomes available, and in some cases for decisions to be revised. Dynamic models are more

compatible with real-world ATFM decision making in the face of unfolding information, but also more complex. Static models can be adapted to dynamic environments by re-running the models at regular intervals.

Disaggregate SATFM models find optimal delays for individual aircraft. These models implicitly assume that a centralized authority determines and controls the delays of individual aircraft. They are frequently based on 0-1 binary integer program making them intractable for a large number of aircraft. Though the disaggregate models that determine efficient decisions for a single airport are typically tractable.

Recent approaches have shifted aggregate flow control. The flow-based models have several advantages over disaggregate ones. First, the flow-based modeling approach is consistent with the Collaborative Decision-Making (CDM) environment. Under CDM, arrival slots are aggregated and allocated to airlines rather than individual aircraft. This allows the airlines to exercise more control and make decisions based on their internal business objectives. These decisions may include flight cancelation or substituting more important flights for less important ones. Second, the flow-based models are compatible with the current in place ration-by-schedule (RBS) principle. In RBS, the arrival slots created by the available arrival capacity at the destination airport are distributed to flights in order of their original, scheduled arrival times at the destination airport. Lastly, from a computation standpoint, the flow-based models involve fewer variables leading to a faster computation time.

**Table 2.2 Dimensions of single destination airport, SATFM models**

| **Revision in decision making** | | **Level of aggregation** | | **Method of modeling the uncertainty in capacity** | |
|---|---|---|---|---|---|
| **Static** | **Dynamic** | **Individual** | **Flow** | **Scenario-based models** | **Markov process** |
| **[21]; [23]** | **[19]; [20]; [22]; [24]; [25]; [26]** | **[19]; [20]; [21]; [22]; [24]; [25]** | **[23]; [26]** | **[19]; [20]; [21]; [22]; [23]; [24]; [26]** | **[25]** |

Finally, we differentiate SATFM models based on their representation of capacity uncertainty. The preponderance of models employs a scenario-based approach. In this case, capacity uncertainty is represented in the form of capacity scenarios and their associated probabilities. Far less common are models that represent the evolution of capacity as a Markov process.

Since it is often difficult to accurately predict the AAR for any significant time into the future, most researchers have formulated GDP models that require a set of probabilistic capacity scenarios to represent capacity uncertainty. A capacity scenario is an arrival capacity profile, which is a time series of airport capacities, typically represented as airport acceptance rates, or AARs. The set of probabilistic capacity scenarios capture the uncertainty in the arrival capacity profile. Each is assigned a particular probability of occurrence. In practice, a scenario will rarely

match the realized capacity profile precisely, but a close resemblance between the realized capacity profile and one or more of the scenarios can be expected. When several scenarios are similar though a certain time period, and then diverge, it is possible to combine them to form a scenario tree, whose branches correspond to the time points when the scenarios become distinct. The branches correspond to distinct individual scenarios or sets of scenarios. The points in time at which decisions need to be made are called nodes. Scenario trees are can capture the unfolding of information about future capacity information over time, and thus support dynamic decision making.

Scenario-based models have been studied in many domains. The literature on scenario generation has developed primarily to support application of stochastic programs in financial portfolio optimization. In portfolio optimization the set of scenarios capture the uncertainty in the random variables that are the returns from various assets including cash, bonds and stocks.

One technique to generate scenarios used in this domain is "moment matching" [**29**]. This technique is useful when the true distribution of the underlying random variable is unknown. The decision maker using his experience or even empirical data provides values of the statistical properties for the distribution of the random variables (like moments, percentiles or covariance matrix). A set of scenarios is constructed that have similar statistical properties as provided by the user. The set of scenarios try to mimic the underlying distribution of the random variable. A commonly used metric of similarity is the sum of the square distance between the values of statistical properties provided by the decision maker and for the set of scenarios. An increase in similarity corresponds to a decrease in the sum of square distances. The number of the scenarios along with the relevant statistical properties needs to be predefined to develop the scenarios.

An important issue is how to evaluate scenario generation methods. It has been persuasively argued ( [**30**] and [**31**]) that modeling the unknown underlying distribution of the random variable is not important as compared to getting "good" decisions from a stochastic optimization model. A "good" decision is one that generates and value for the objective function close to what would be obtained using the true distribution of the random variable. Therefore, scenarios the yield such decisions are desired.

When the distribution of the underlying random variable is known then sampling based methods can be used to generate the scenarios [**30**] and [**32**]. The basic idea is to sample from the underlying distribution at every time epoch given the value of the random variable in the previous time epoch. References propose sampling based methods to generate the scenario trees. Given the conditional probability density function (PDF), $(x_{t+1}|x_t)$ where $x_t$ is the value of the random variable at $t$, the value and the associated probability at $t+1$ can be determined by sampling from this conditional probability density function (PDF). Since sampling is performed at every time epoch, the scenarios grow exponentially with the increase in the length of planning horizon and with the increase in the number of scenarios. This type of scenario generation methodology requires predefining the number of scenarios.

A method of generating the scenario tree is to generate individual scenario paths and then bind the common scenario paths to form a tree [**33**].

The above literature on scenario generation is primarily focused on portfolio optimization and offers a limited scope in developing probabilistic capacity scenarios for arrival capacity profile. This can be attributed to several reasons. Firstly, predefining the structure of the scenarios is suitable towards financial applications, as the nodes may represent distinct decision points like an expiration of an option. Secondly in finance, the returns over time do not show any exploitable historical trends whereas, the arrival capacity exhibits seasonal and daily trends. Therefore probabilistic capacity scenarios can be determined from historical arrival capacity data that exploit the historical trends. In finance, the returns can theoretically take any value and therefore simulation based scenario generation approaches are suitable. Whereas, the arrival capacities are predetermined by the FAA, that are based on the weather conditions and runway configurations. Therefore the arrival capacity scenarios have a smaller set of discrete, integer values. The scenarios can be generated without warranting a simulation. Lastly, having a fixed number of scenarios in finance might allude to a fixed number of movements in the returns of an asset. In aviation, the number of scenarios may vary on a daily basis. Therefore a fixed number of scenarios might add redundancy or under represent the uncertainty in capacity depending on the day. Reference [**25**] confirms the some of the above shortcomings and stresses that recourse actions in an SATFM model depend on the structure of the scenarios. Therefore, the structure of the scenario tree is of paramount importance in SATFM.

A potential pitfall for scenario based methods is it is rare that a realized capacity profile is matched precisely with a scenario. An illustrative example is shown in Figure 2.8. It shows the six probabilistic capacity scenarios developed for SFO in reference [**34**] with a realized arrival capacity profile.

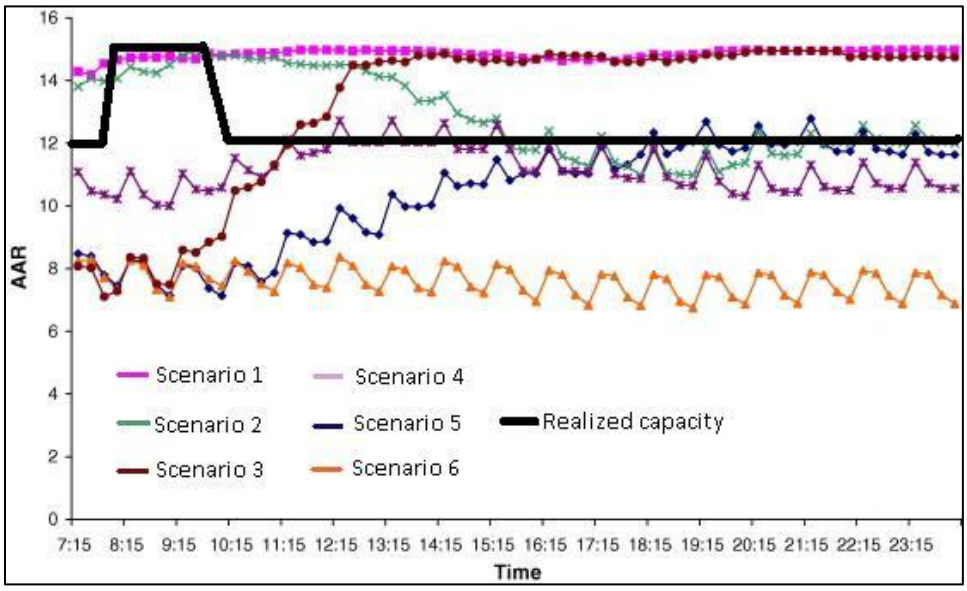


**Figure 2.8 Scenarios and realized capacity**

The second method of capturing the uncertainty in the arrival capacity is by modeling the evolution of capacity as a Markov process. This assumes that the capacity at time $T + 1$ depends on the realized capacity between times $[T - \delta, T](\delta \geq 0)$ and independent of earlier times. In a GDP the decisions are made sequentially for every time period are based on evolving capacity

information which can be captured by a Markov process. The efficient strategic decisions for every period can then be determined using a dynamic program. This method of modeling the uncertainty in capacity using a Markov process and a dynamic program is also referred as a "scenario free" approach. The scenario free approach is investigated in a greater detail by reference [**25**]. A potential pitfall of the "scenario free" approach is the curse of dimensionality associated with the dynamic program.

This research aims to support the scenario-based approach. This is because scenario-based models are more common in the air traffic management literature, more tractable, and more directly applicable to current air traffic management practices.

### 2.3.1 Determining the arrival capacity from historical data

Methods for developing probabilistic capacity scenarios have been reported in the ATFM literature. One approach [**34**] is based historical arrival capacity profiles. Scenarios are the averages of clusters of similar profiles that are identified through K-means clustering on the historical profiles. This approach does not take into account any weather forecast or other day-of-operation information. Reference [**35**] presents a GDP model based on the SFO Marine Stratus Initiative (STRATUS) forecast. The time of fog burn-off—and consequent capacity increase—is modeled as a discrete random variable with the probability distribution obtained from historical burn-off time combined with a point estimate for the time for a given day derived STRATUS. The capacity profile for a given burn-off time is based on the assumption that at fog burn-off the landing capacity of SFO increases abruptly from its normal instrument metrological condition (IMC) to its normal visual metrological condition (VMC) level. Other sources of capacity variability, while evident in the data as shown in Chapter 3, are not in [**35**]. Furthermore, the research uses the STRATUS forecast and is thus applicable only to SFO. There are no known methods that develop probabilistic capacity scenarios from generic weather forecasts available at all airports. A method for determining a single scenario from a forecast has, however, been developed [27]. This involves matching the realized historical weather in a period with the capacity of the airport in that period, and then using this relationship to predict the AAR from the Terminal Aerodrome Forecast (TAF), and also from the Meteorological Aviation Report (METAR), for every period. This approach does not take into account the uncertainty concerning the TAF itself or in the relationship between realized weather and airport capacity.

## 2.4 Static Stochastic Ground Delay Model

This research proposes to use a static, flow based aggregate model that requires probabilistic capacity scenarios to model the uncertainty in the arrival capacity to make flow management decisions. To this end this research chooses the Static Stochastic Ground Delay Model (SSGDM) [**23**]. The SSGDM is employed to evaluate the performance of the various scenario generation methodologies in a real world setting. The SSGDM determines an optimal ground delay strategy corresponding to different scenario generation techniques thereby enabling comparisons between them.

As the name suggests this model is a static model: ground delay decisions are made at a single point in time and are not revised as new information about the capacity is obtained. An input to this model is the set of probabilistic capacity scenarios and therefore different methodologies of scenario generation can be compared.

SSGDM determines optimal ground delay decisions for flights bound for a single destination airport. While the model ignores the capacity uncertainty in the en-route airspace, it employs probabilistic scenarios representing the uncertainty in the arrival capacity at the destination airport as inputs. The model requires the original scheduled demand profile for flights bound for the destination airport along with the relative costs of delays taken in the air or on the ground. It minimizes the sum of ground delay costs and expected airborne delay costs. The output from the model is the arrival rate of flights at the destination airport, called the Planned Airport Arrival Rate (PAAR), the number of the flights subjected to ground delays in each period, and the number of flights subjected to air delay in each period under a particular capacity scenario.

The SSGDM is given by equation (2.1) to equation (2.4).

$$Min\left[\sum_{t=1}^{T} c_g \times G(t) + \sum_{k=1}^{n}\sum_{t=1}^{T} c_a \times W(S_k,t)P(S_k)\right] \quad 2.1$$

The constraints to the above objective function are given below

$$R(t) - G(t-1) + G(t) = D(t)$$
$$(t \in 1,2,..,T+1; G(0) = G(T+1) = 0) \quad 2.2$$

$$-W(S_k,t-1) + W(S_k,t) - R(t) \geq -M(S_k,t)$$
$$(t \in 1,2,..,T+1; -W(S_k,0) = -W(S_k,T+1) = 0, \quad 2.3$$
$$k \in 1,2,..,n)$$

$$R(t), W(S_k,t), G(t) \in Z_+ \ (t \in 1,2,..,T+1, k \in 1,2,..,n) \quad 2.4$$

Where:

t is time period;

Tis the total number time periods

$S_k$ is the $k^{th}$capacity scenario;

$P(S_k)$ is the probability of the $k^{th}$ scenario ;

n is the number of scenarios;

G(t) is number of flights that are ground delayed in period t;

$W(S_k, t)$ is number of flights that are air delayed in period $t$ under capacity scenario $S_k$;

$R(t)$ is the arrival rate of flights referred to as the planned airport acceptance rate (PAAR) at time $t$ ;

$M(S_k, t)$ is the arrival capacity under scenario $S_k$ at time $t$;

$D(t)$ is arrival demand in period $t$;

$c_a$ is unit cost of airborne delay (= 3);

$c_g$ is unit cost of ground delay (=1)

The objective function to be minimized, equation (2.1), is the sum of the deterministic ground delay costs and the expected air delay costs. Equation (2.2) is a flow conservation constraint for flights bound for the destination from all the origin airports. The demand at period $t$, $D(t)$, plus the backlogged flights ground held in period $t-1$, $G(t-1)$, must either land, and thus count toward $R(t)$, or be put in a queue, contributing to $G(t)$. Equation (2.3) is a queuing constraint at the destination airport. Under capacity scenario $S_k$ at time period $t$, the flights accruing air delay in $t-1$, $W(S_k, t-1)$, plus the flights assigned arrival times in time period $t$, $R(t)$, either land or are air delayed for another period, contributing to $W(S_k, t)$. The inequality is required as the total demand might be less than the available capacity. Equation (2.4) ensures that $R(t)$, $W(S_k, t)$ and $G(t)$ are real positive integers.

The decision variables are the arrival rates of flights at the destination airport, $R(t)$, the number of flights that are subjected to ground holding, $G(t)$, and the number of flights subjected to air delay under the $k^{th}$ scenario , $W(S_k, t)$ in a period $t$.

## 2.5 Contributions to the literature

The goal of this research is to improve the service provider's (Federal Aviation Administration, FAA) strategic decision making while effectively utilizing the weather forecasts. In particular, this research will focus on using weather forecasts to better plan a GDP thereby reducing the delay and the associated delay costs.

We showed earlier that the airport arrival capacity or the AAR is strongly influenced by the weather in the vicinity of the airport. Therefore to predict the AAR we require an airport specific weather forecast. There are two types of uncertainty associated with capacity prediction from weather forecast that need to be addressed. First, it is known that weather forecasts are seldom accurate in predicting the actual weather conditions. Therefore there is an uncertainty associated with the weather forecasts. Second, the weather forecast is not a direct predictor of arrival capacity. Even if the forecast is of high fidelity, there is still an inherent uncertainty in determining capacity values from the forecast conditions. This research contributes by developing methodologies that capture the uncertainty in capacity from the forecast.

Our research would assist the service providers in planning a GDP better. Currently, air traffic managers familiar with a NAS resource use the current weather, weather forecast and the future demand projections to make strategic decisions while balancing the preferences of the operators. There exists no formal mechanism with which the air traffic managers translate the weather forecast into capacity values. The resulting decisions tend to be conservative, resulting in excess delays and associated delays costs. To illustrate with an example, the ground delay program at San Francisco International airport typically lasts two hours longer than required [35]. This happens because air traffic managers are unsure about the improving weather conditions and are thus hesitant to call a higher rate as it may compromise safety. With our methodology the air traffic manager would be able to translate the weather forecast into probabilistic capacity scenarios. These scenarios would assist them in planning ground delay operations. Furthermore, as the day-of-operation progresses, the scenarios can be updated with the change in the weather forecast allowing revisions in the ground delay programs.

Using our probabilistic capacity scenarios as inputs, the existing GDP models can be implemented for real operation planning. Strategic decisions, like arrival rates in a ground delay program (GDP), rely on the weather forecast and are made more than two hours before their implementation, when there is an uncertainty in the future arrival capacity. The GDP models developed in the literature require the uncertainty in the future arrival capacity as inputs to determine optimal arrival rates. This uncertainty in the future arrival capacity is captured either by a set of scenarios or modeled as a Markov process. It is assumed that the scenarios or the transition probabilities can be determined from weather forecasts or by the expertise of air traffic managers. The optimal arrival rates determined from these models are based on notational probabilistic capacity scenarios or assumed transition probabilities. This hinders the real world application of the GDP models. By focusing on the more prevalent input, probabilistic capacity scenarios, this research develops methodologies to generate the day-of-operation probabilistic capacity scenarios from the weather forecast. Two of the methodologies utilize the day-of-operation Terminal Aerodrome Forecast (TAF) along with the past TAF and realized capacity to develop unique day-of-operation probabilistic capacity scenarios. Such scenarios, when used in conjunction with appropriate GDP planning models, could lead to better GDPs, with lower costs by avoiding excessive ground delays or airborne delay.

One of the TAF based methodologies requires three airport specific input parameters. These three input parameters determine the number and the probability of probabilistic capacity scenarios. It is important to tune these parameters, such that the resulting scenarios give the lowest possible average realized costs, thereby maximizing the performance of the scenarios. The combination of values for the three input parameters which result in the lowest average realize cost is referred as the "optimal design" for an airport (by "optimal design" we don't mean the physical design of the airport but rather the parameter values which minimize the realized costs). This research uses state-of-the-art optimizing techniques to determine the "optimal design". This research finds the "optimal design" to vary among airports, in some cases greatly. It is shown that scenarios developed from weather forecasts control the cost of delays on days that have a greater capacity-demand imbalance as compared to scenarios developed without weather forecasts.

We provide a platform to measure the efficacy of the scenarios in a real world setting based on several days in the past. Researchers have traditionally compared the performance of their

models by comparing the expected delay costs (the objective function) from the stochastic GDP models to the cost under the perfect information. Here we present a platform that measures the realized cost based on the planned arrival rate output from the SSGDM and the realized capacity profile for each of the historical day. The realized cost of delay (TCD) is the sum of ground delay costs and the realized air delay costs. The average of the realized cost of delay over all the days in the past is called the average realized cost of delay($\overline{\text{TCD}}$). ($\overline{\text{TCD}}$) is our metric that measures the efficacy of the scenarios. We also quantify the benefit of using weather forecasts in decision making using our platform. We use our platform to compare $\overline{\text{TCD}}$ determined from the scenarios generated by the weather forecasts with $\overline{\text{TCD}}$ determined from scenarios generated without forecast information. This research also compares the costs determined from the scenarios with the costs under planning operations based on perfect information. We will apply and assess the performance of the scenarios for US airports: Boston Logan International Airport (BOS), Chicago O'Hare Airport (ORD), San Francisco International Airport (SFO) and Los Angeles International Airport (LAX). We also present a methodology to determine scenarios from STRATUS, the specialized forecast product predicting the time of fog burn-off exclusively for SFO. We use the STRATUS forecast to develop an exclusive methodology for SFO to generate probabilistic capacity scenarios. We will show that scenarios generated from weather forecasts on average reduce the realized costs of delays for every airport.

Below we present a table that introduces the methodology names, a brief description, the data required by the methodology, the key statistical techniques for generating the scenarios, and the sections where they appear in this thesis.

**Table 2.3 Overview of the scenario generation methodologies**

| Methodology name | Description | Data | Statistical techniques | Section |
|---|---|---|---|---|
| TAF Clustering | Determining representative profiles for days having similar weather forecast | Terminal Aerodrome Forecast, realized capacity profiles | Principal component analysis, K-means clustering, Silhouette value | 3.3.2 |
| Fog burn-off time clustering (exclusively for SFO) | Determining representative profiles for days having similar fog burn-off time | STRATUS forecast, realized capacity profiles | Cumulative density function, binning of data | 3.3.3 |
| DTW Scenarios | Scenarios are capacity profiles of days having similar weather forecast. | Terminal Aerodrome Forecast, realized capacity profiles | Dynamic time warping, Stochastic response surface methodology, design-of-experiments | 4.2 |
| Naïve Clustering (reference case) | Determining representative profiles for days | Realized capacity profiles | K-means clustering, Silhouette value | 3.4.1 |

| | | | | |
|---|---|---|---|---|
| | having similar arrival capacity profile. No weather forecast. | | | |
| Perfect information (reference case) | Precise knowledge about the capacity profile | Realized capacity profile | - | 3.4.2 |

# Chapter 3 Scenario generation from weather forecasts

## 3.1 Introduction

This chapter discusses methodologies to generate probabilistic capacity scenarios. Arrival capacity is strongly influenced by the weather in the vicinity of the airport and thus AAR prediction necessitates a terminal weather forecast. This results in uncertainty about the arrival capacity. However, as argued above, it may be possible to characterize and manage this uncertainty by developing capacity scenarios from weather forecasts.

This chapter first discusses the data that will be used to generate the probabilistic capacity scenarios. These data include historical arrival capacity data, as well as data from the Terminal Aerodrome Forecast (TAF) and the San Francisco Marine Status Initiative (STRATUS) forecast. The TAF is issued for all major airports where as the STRATUS forecast is exclusive to SFO. We then introduce the notation for scenario development used in the subsequent sections. This chapter then proceeds to presents two methodologies for generating the probabilistic capacity scenarios from a weather forecast. The first methodology uses the TAF and second methodology requires the STRATUS forecast. Since, the first methodology uses the TAF it can therefore be applied to all major airports. The second methodology requires the STRATUS forecast and therefore its application is limited to SFO. Next, we define scenarios for two reference cases, one featuring perfect information and the other not having access to a day-of-operation weather forecast, that enable us to assess the forecast-based methodologies. The chapter concludes by presenting a procedure, based on determining the realized costs of delays using the scenarios in decision making on historical days. This procedure will be used to assess the various scenario generation methods described in the earlier sections.

All the methodologies except the one requiring the STRATUS forecast can be applied to any major airport. The scenarios that are obtained for different airports as the outputs of these methodologies are presented in Chapter 5.

## 3.2 Data for generating probabilistic capacity scenarios

This research develops probabilistic capacity scenarios that represent the uncertainty in the arrival capacity profile from 7:00 am to 10:00 pm, as the bulk of the traffic is in this time period. The scenarios are developed using the historical capacity profiles and issued weather forecasts for the months of May to September from 2004 to 2006.

### 3.2.1 Airport arrival capacity data

The maximum number of arriving aircraft that the airport can accommodate in a given period is called the Airport Acceptance Rate (AAR). The AAR is determined from the guidelines issued by the FAA that include the metrological conditions, the separation and the runway configuration at the airport. The airport control tower determines the AAR for a period based on the prevailing

metrological conditions. The terminal radar approach control (TRACON) meters the air traffic into the terminal airspace based on the AAR.

The Aviation System Performance Metrics (ASPM) website [**10**] maintains records of the AAR for the major airports for historical days starting from 2000. The AAR reported on the Aviation System Performance Metrics (ASPM) website is available in hourly or quarter- hourly formats. In either case, only integer values are allowed, leading to periodic oscillations in some cases. For example, the quarter-hour decomposition when the AAR is 60 arrivals per hour is 15, 15, 15 and 15 arrivals in each quarter-hour, while an AAR for 45 arrivals per hour is decomposed as 12,11,11 and 11 arrivals in each quarter-hour, and an AAR of 30 arrivals per hour is reported as 8, 7, 8 and 7 arrivals per quarter-hour. This decomposition causes oscillations in the AAR reported for quarter-hour period. Figure 3.1 depicts the AAR capacities for July 3, 2007 at SFO. The oscillation can be seen in the quarter-hour profile when the arrival capacity is 30 arrivals per hour. In this research, we use the quarter-hourly format.

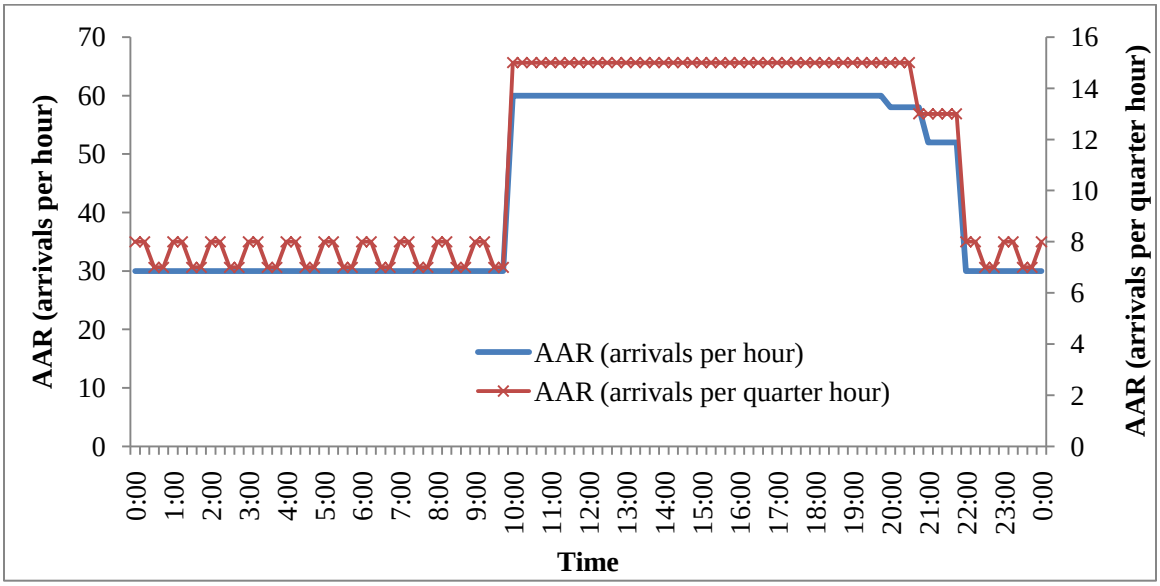


**Figure 3.1 Arrival capacity profile hourly and quarter-hourly for SFO on 07/03/2007**

An arrival capacity profile represents the arrival capacity as measured by the airport acceptance rate (AAR) by quarter-hour from 7:00am to 10:00 pm, which is equivalent to 60 quarter-hour periods. We define $A^d \in I^{60}$ as the quarter-hour capacity profile for the $d^{th}$ day.

### 3.2.2 Terminal Aerodrome Forecast

The Terminal Aerodrome Forecast (TAF) is a weather forecast issued for every major airport forecasting meteorological conditions for the subsequent 24 hours. The TAF is issued four times a day at six hour intervals by the National Oceanic and Atmospheric Administration (NOAA). It predicts meteorological conditions for five statute miles (SM) around an airport [**36**]. It contains the four letter International Civil Aviation Organization (ICAO) airport identifier, the date and time when the forecast was issued along with the duration within which the forecast is valid. The TAF predicts the wind speed, wind direction (reference true north), visibility, and ceiling of four

different cloud types: broken, scattered, overcast, and few. We decompose the wind speed into two perpendicular constituents representing wind blowing from the north and from the east using the angle of wind reported in the TAF. The direction of the wind blowing from north to south and from east to west is considered positive. When the TAF forecasts clear skies, we assumed the cloud ceiling to be present at 250,000 feet for all the four clouds types. Further, when certain cloud ceiling heights are absent in the forecast for a quarter-hour period in the TAF, they are also assumed to be present at 250,000 feet. The TAF may also predict qualitative descriptors like rain, fog and mist for the airport. If the TAF forecasts a thunderstorms occurring close to the airport, it may also contain the probability of its occurrence. The TAF for historical days for different airports is available for download from the National Oceanic and Atmospheric Administration (NOAA) website [**37**]. This research uses the wind speed from north, wind speed from east, visibility and the ceiling of the four cloud types, a total of seven forecast attributes for generating scenarios. Qualitative variables like rain and hail are ignored because they are infrequently represented in the historical TAF data.

The TAF issued between 4am and 6am, local time for the different airports, is used for developing the probabilistic capacity scenarios. In practice, earlier TAF forecasts must also be used to make traffic flow management decisions before flights have departed for that airport.

The predictions for the TAF forecast are decomposed into a quarter-hour format. This decomposition is required to match the forecasts with the quarter-hour arrival capacity. Therefore, each quarter-hour period is associated with seven metrological attributes. A sample TAF issued for SFO on 6-17-2004 along with its explanation is shown below.

KSFO 171120Z 171212 19008KT P6SM SCT015
TEMPO 1418 BKN015
FM2000 23015KT P6SM SKC
FM0400 24010KT P6SM SCT015
FM0600 25008KT P6SM BKN012

This TAF was issued for SFO on the 17$^{th}$ of the month at 1120Z time (Z means Zulu time, GMT that is PDT+7 hours). This forecast is valid from 1200Z on the 17$^{th}$ to 1200Z on the 18$^{th}$. For the periods between 1200Z and 2000Z, the wind speed and direction is predicted to be 8kt and 190 degrees respectively. The visibility is greater than 6 SM, with scattered clouds at 1500 feet and the other cloud ceiling heights are absent. Between 1400Z and 1800Z broken clouds at 1500 feet are forecast. For the periods between 2000Z on the 17$^{th}$ and 0400Z on the 18$^{th}$ the wind is forecast at 15Knots blowing from 230 degrees and visibility is greater than 6SM with clear sky conditions. From 0400Z to 0600Z, the wind is predicted to be 10kt from 240 degrees with visibility greater than 6SM and scattered clouds at 1500 feet, while the other cloud ceiling heights are absent. From 0600Z to 1200Z on the 18$^{th}$, the wind speed and direction is 8Kt and 250 degrees respectively. The visibility is in excess of 6SM with broken clouds at 1200 feet during this period.

We employ two representations of the TAF, in order to support scenario generation methods presented in the subsequent sections. The first is a column vector format while the second represents the TAF as a multidimensional time series. A multidimensional time series describes the TAF by its seven attributes called dimensions at multiple, discrete time periods.

$T^d$ is the TAF for the $d^{th}$ day, in a column vector format, length 60 (quarter-hour periods) × 7(attributes per time period)=420 units.

$\vec{T}^d$ is the TAF for the d$^{th}$ day in a multidimensional time series format, dimension 60× 7 units.

$\vec{T}^{d,r}$ is the TAF forecast for the $d^{th}$ day for the quarter-hour period $r$, representing the wind speed from north, wind speed from east, visibility, ceiling of few, broken, overcast and scattered clouds, dimension 1× 7 units.

### 3.2.3 San Francisco Marine Stratus Initiative (STRATUS)

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 simultaneous landings on parallel runways 28L and 28R, reducing the arrival capacity from 60 arrivals to 30 arrivals per hour. If the fog persists into the hours when heavy traffic is scheduled, it creates a demand-capacity imbalance. On such days a Ground Delay Program is initiated at SFO. The primary purpose of the STRATUS forecast system is to predict the fog burn-off time.

The San Francisco Marine Stratus Initiative (STRATUS) forecast system is weather forecast product for San Francisco International Airport (SFO) developed by Massachusetts Institute of Technology-Lincoln Labs (MIT-LL). It predicts the time when the fog dissipates from the SFO terminal area, also called the burn-off time. STRATUS also predicts the probability of burn-off before 10:00am, 11:00am, 12:00pm and 1:00pm PDT (the STRATUS reports the probabilities at 17Z, 18Z, 19Z and 20Z). The predictions for the burn-off time are updated hourly from 2am to 11am based on evolving information and weather conditions. The burn-off time is determined using a weighted average of the burn-off time predicted by the four individual models: local statistical model (LSM), regional statistical model (RSM), satellite statistical model (SSM) and an atmospheric boundary layer physics model (in French, Couche Brouillard Eau Liquide, COBEL). The inputs to the models vary significantly. The individual weights are determined from the historical accuracy of individual models. The probabilities of burn-off by various times are based on correlating the actual burn-off time of the historical days with the forecast burn-off time [**38**].

The STRATUS forecast for the historical days was made available from NASA Ames Research Center. This data contains the days when the STRATUS forecast was issued for SFO i.e. the days when fog was forecast for SFO. The data includes the predicted burn-off time, as well as the probability of burn-off before 10 am, 11 am and 12 pm. NASA Ames Research Center also maintains a repository where the actual fog burn-off time is recorded, which we denote as τ. In total there were 183 days between 2004 and 2006 when the fog was observed at SFO. Because STRATUS is available only on days when low lying clouds are forecast for the airport, the number of days when the STRATUS forecast is available is smaller than number of days when TAF is available.

We use the forecast issued at 8am as it's the first forecast of the day using predictions from the SSM. In practice, earlier forecasts must also be used to make traffic flow management decisions before flights have departed for SFO. Here we'll assume for simplicity that all the decisions can be based on the more reliable 8:00am forecast.

Figure 3.2 shows a scatter plot of the actual and predicted fog burn-off time along with a trend line.

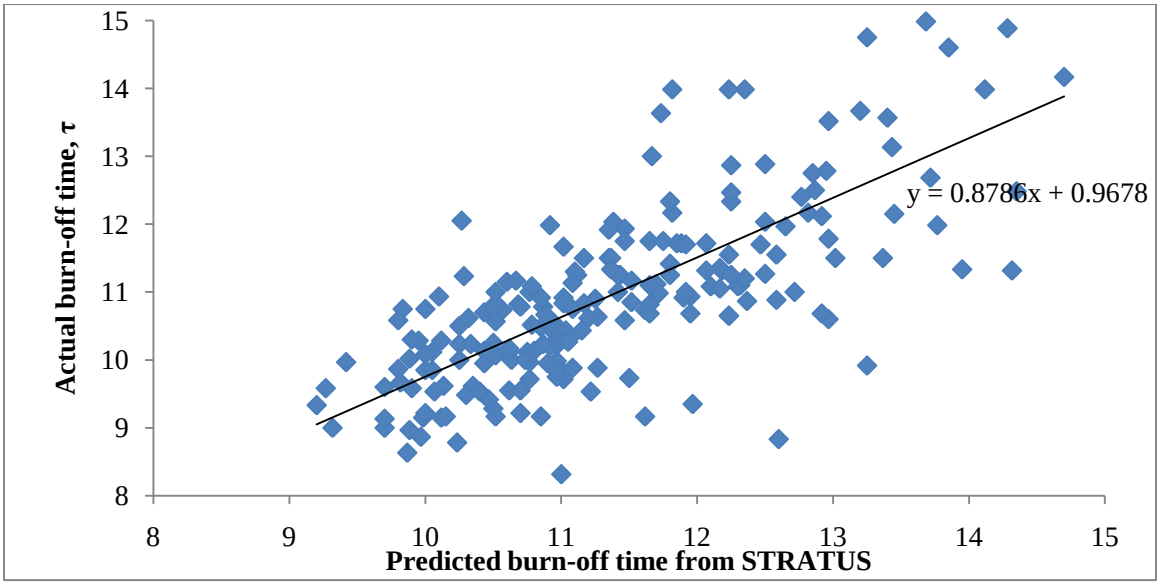


Figure 3.2 Predicted versus actual burn-off time from STRATUS for SFO

Figure 3.3 shows the cumulative distribution function (CDF) of actual burn-off time at SFO. It can be noticed that for 75% of the days, the fog burn-off time is between 9:30 am and 11:30 am.

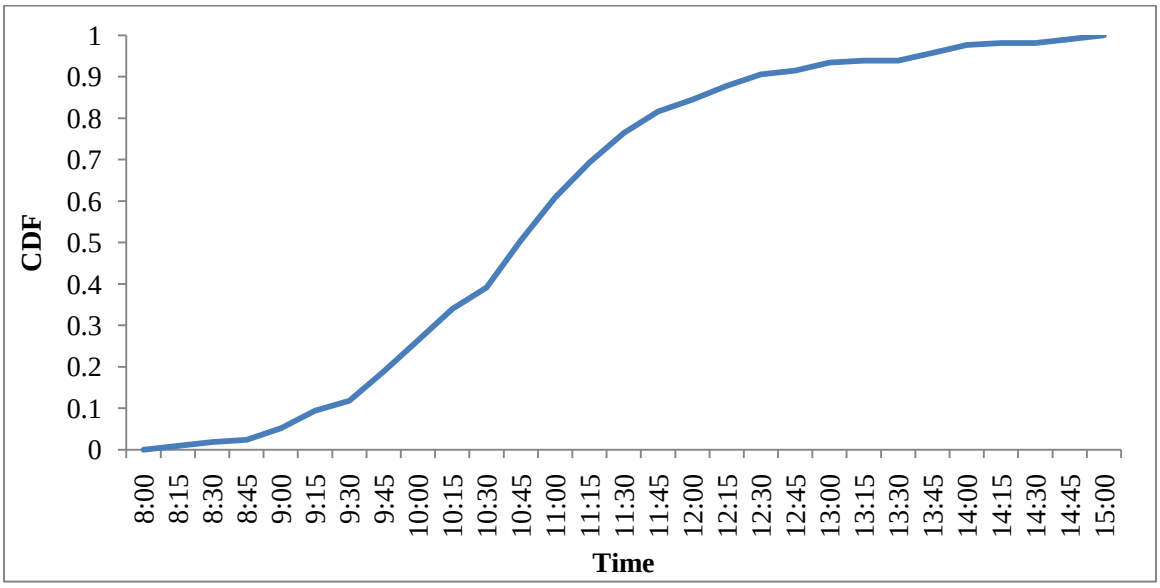


Figure 3.3 Distribution of days according to actual burn-off time

Table 3.1 shows an example the STRATUS forecast, generated at 8 am for 6-17-2004. For this day, STRATUS predicted the burn-off time as 11:31am. STRATUS also predicted a 5% probability of the fog burn-off time before 10am, a 30% probability of the fog burn-off time before 11am and a 65% probability of fog burn-off before noon. The actual fog burn-off time, $\tau$, recorded by NASA Ames was 10:51am, therefore STRATUS over predicted the burn-off time by 40 minutes.

**Table 3.1 Sample STRATUS forecast**

| Date | Predicted burn-off time | P(10) | P(11) | P(12) | Actual burn-off ($\tau$) |
|---|---|---|---|---|---|
| 6/17/2004 | 11:31 | 0.05 | 0.3 | 0.65 | 10:51 |

## 3.3 Scenario generation using weather forecasts

This section presents the scenario generation methodologies. First, the notations used throughout the chapter are introduced. This is followed by the scenario generation methodology using the TAF and then the specific scenario generation methodology for SFO using the STRATUS forecast.

### 3.3.1 Notation

$\tau$ is the actual fog burn-off time
$F_\tau(t)$ is the probability of burn-off before time $t$
$D$ is the total number of historical days for an airport
$C_k$ is the $k^{th}$ cluster
$|C_k|$ is the number of days in the $k^{th}$ cluster, $C_k$
$A^i$ is the Arrival capacity profile for the $i^{th}$ day, length 60 units
$A_k^i$ is the AAR capacity profile of the $i^{th}$ day within the $k^{th}$ cluster, $C_k$
$T^i$ is the TAF vector for the $i^{th}$ day, length 420 units
$T_k^i$ is the TAF vectors of the $i^{th}$ day within the $k^{th}$ cluster, $C_k$
[T] $= [T^1 T^2 .. T^D]$ , a TAF matrix comprising of individual TAF vectors, dimension $420 \times 446$
$\vec{T}^d$ is a multidimensional time series of TAF for the $d^{th}$ day, dimension $60 \times 7$
$\vec{T}^{d,r}$ is the TAF attributes for a quarter-hour period $r$ for the $d^{th}$ day, dimension $1 \times 7$
$S_k$ is the Probabilistic capacity scenario for the $k^{th}$ cluster, $C_k$
$P(S_k)$ is the probability of the probabilistic capacity scenario $S_k$
$[\,.\,]$ is the nearest integer

### 3.3.2 TAF Clustering

This scenario generation methodology can be applied to all major airports that are issued a TAF.

This methodology generates probabilistic capacity scenarios for days having similar TAFs. Having decomposed the TAF for a given day into quarter-hour periods, the TAF for any historical day, $d$, can be represented by a column vector, $T^d$ of 420 attributes (60 quarter-hour periods × 7 attributes per quarter-hour = 420 attributes). We seek to identify groups of days that have similar TAF vectors. Toward this end, we apply clustering. Such a high number of attributes, many of which are highly correlated (for example, attributes in consecutive time periods) makes the direct application of clustering techniques inappropriate [**39**]. Therefore, we first reduce the number of attributes required to represent $T^d$ to a smaller, uncorrelated set of variables. We then find groups of days having similar TAFs (now represented by a smaller number of attributes) through K-means clustering. The magnitude of variation between the original attributes should also be captured by the new, reduced number of attributes. This can be accomplished using a statistical technique called Principal Component Analysis (PCA). We perform PCA prior to clustering for two reasons: Firstly, PCA reduces the number of attributes required in expressing the data thereby making clustering more tractable. Secondly, it has been proved in that performing PCA prior to K-means clustering increases the accuracy of determining optimal clusters [**40**].

Define $[\mathrm{T}] = [T^1\ T^2\ ..T^D]$ to be $420 \times \mathrm{D}$ matrix comprising of individual TAF vectors. We performed PCA on this matrix. PCA reduces the dimensionality of the data by representing correlated attributes with a smaller number of uncorrelated attributes called principal components. The first principal component represents the direction of maximum variation in the data and each succeeding component represents the direction of the maximum remaining variation in the data. PCA removes correlation between the forecast attributes [**41**]. As a standard preprocessing technique, we first standardize the [T] matrix i.e. the mean and the standard deviation is 0 and 1 respectively for each attribute. Equation (3.1) through equation (3.5) describes the PCA on the data set.

$$[\mathrm{M}] = {}^{[\mathrm{T}][\mathrm{T}]^{\mathrm{t}}}\!/_{\mathrm{D}-1} \qquad 3.1$$

Where [M] is defined as the empirical correlation matrix and $[\mathrm{T}]^{\mathrm{t}}$ is the transpose of matrix [T]. The eigenvalues, $\lambda$, of the empirical correlation matrix, [M], determine the magnitude of variation in the data and the eigenvectors, X, the direction of variation.

$$[\mathrm{M}]X = \lambda X \qquad 3.2$$

If we assume the matrix is full rank then there are 420 eigenvalues $(\lambda_1, \lambda_2, .., \lambda_{420})$ and corresponding eigenvectors $(X_1, X_2, .., X_{420)})$. A greater value of $\lambda_i$ implies a greater variation in the direction of $X_i$.

In equation (3.3), the eigenvalues are sorted in a descending manner such that $\lambda_{<i>}$ is the $i^{th}$ greatest eigenvector and $X_{<i>}$ is the corresponding eigenvector.

$$\lambda_{<1>} > \lambda_{<2>} > \lambda_{<3>} > . . > \lambda_{<420>} \quad 3.3$$

A standard technique is to capture at least 90% variability in the data, in which case the number, $n$, of eigenvalues and eigenvectors required is given by equation (3.4).

$$n = \min s\,; subject\ to\ \frac{\sum_{t=1}^{s} \lambda_{<t>}}{\sum_{p=1}^{420} \lambda_{<p>}} \geq 0.9 \quad 3.4$$

We then define a matrix $[\mathrm{W}] = [X_{<1>} X_{<2>} .. X_{<n>}]$ comprising of the n eigenvectors corresponding to the $n$ greatest eigenvalues. The matrix [W] represents the directions that capture at least 90% of the variation in the original TAF attributes. Finally we reduce the attributes of the TAF vectors according to equation (3.5).

$$\begin{aligned} \left[\tilde{\mathrm{T}}\right] &= [\mathrm{W}]^{\mathrm{t}} \times [\mathrm{T}] \\ \left[\tilde{\mathrm{T}}\right]_{n \times D} &= [\tilde{T}^1\ \tilde{T}^2\ .. \tilde{T}^D] \end{aligned} \quad 3.5$$

Where $\tilde{T}^d$ is the reduced n-dimensional TAF vector for the $d^{th}$ day.

We next found groups of days that have similar weather forecasts based on the reduced TAF vectors. This is achieved by performing a K-means clustering on the reduced TAF vectors $\tilde{T}^1, \tilde{T}^2, .., \tilde{T}^D$. The measure of similarity between two days is gauged by the Euclidean norm of the difference between the two reduced TAF vectors. A smaller Euclidean norm indicates a greater similarity between the two days. Therefore, a K-means clustering operation with $L$ predefined clusters partitions $\tilde{T}^1,\ \tilde{T}^2, .., \tilde{T}^D$ into $L$ distinct clusters. Each cluster, $C_l$, contains $|C_l|$ days ($l \in 1,2,..L$).

Determining the optimal number of clusters, $L^*$, is an open problem for which several ad-hoc procedures have been developed. A previous research effort provides an algorithm involving the pseudo-F statistic, combined with a required minimum number of observations within each cluster, to determine the number of clusters [**34**]. The pseudo- F statistic measures the compactness of a cluster with respect to other clusters and reports an average value over all clusters. It calculates a ratio of the mean sum of squares between different clusters to the mean sum of squares within a cluster. Higher pseudo F-values indicate tight clustering and imply that the observations are well separated or better clustered. It has been reported that the pseudo- F statistic works better with data that follows approximately a multivariate normal distribution.

Another technique for determining the number of clusters employs "silhouette value." For each observation this technique determines a "silhouette value", which is a function of two similarity metrics: First, the average similarity between the observation and cluster in which the observation is classified and second, the minimum average similarity between the observation and other clusters. Figure 3.4 shows three clusters, $A, B$ and $C$ with 3,3 and 4 observations within

the clusters respectively. The silhouette value for observation $i$ in cluster $A$ is given by equation (3.6).

$$s(i,A) = \frac{b(i,A) - a(i,A)}{\max(a(i), b(i))} \qquad 3.6$$

Where,

s(i, A) is the silhouette value for observation i in cluster A

T(k) is the attributes of observation k

|Y| is number of observations in cluster Y

a(i, A) is the average similarity (based on the difference of the Euclidean norm) of observation i to all other observations in cluster A.

$$a(i,A) = \frac{\sum_{k \in A \setminus i} \left| |T(i) - T(k)| \right|^2}{|A| - 1}$$

$d(i,Y)$ is the average similarity (based on the difference of the Euclidean norm) of observation $i$ to all observations in cluster $Y$ ($Y \neq A$).

$$d(i,Y) = \frac{\sum_{k \in Y} \left| |T(i) - T(k)| \right|^2}{|Y| - 1} \; ; \forall Y \in \{B, C\}$$

$$b(i,A) = \min_Y d(i,Y) \; ; Y \in \{B, C\}$$

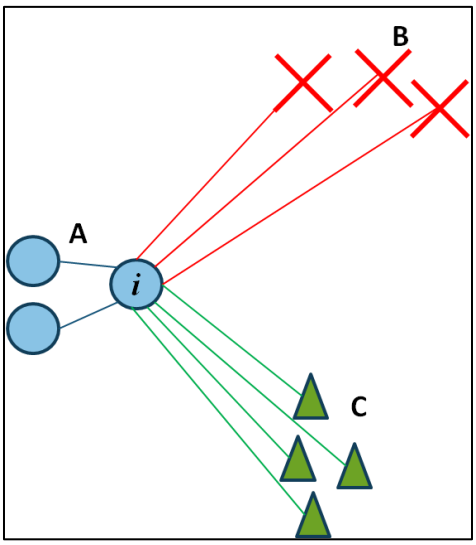


Figure 3.4 An illustration of silhouette value calculation

The silhouette value varies between -1 and 1 with a value close to 1 indicating of well-separated cluster [**42**]. For a silhouette value to be close to 1, $b(i)$ has to be large indicating that observation $i$ is distinct from observations in other clusters or $a(i)$ has to be small indicating that the observation is similar to observations within the same cluster.

As an illustration, the silhouette value plots along with the average silhouette values for two and three clusters for SFO are shown in Figure 3.5. The observations are better separated when there are two clusters because of the higher average silhouette value. When the number of clusters is three it results in a lower average silhouette value.

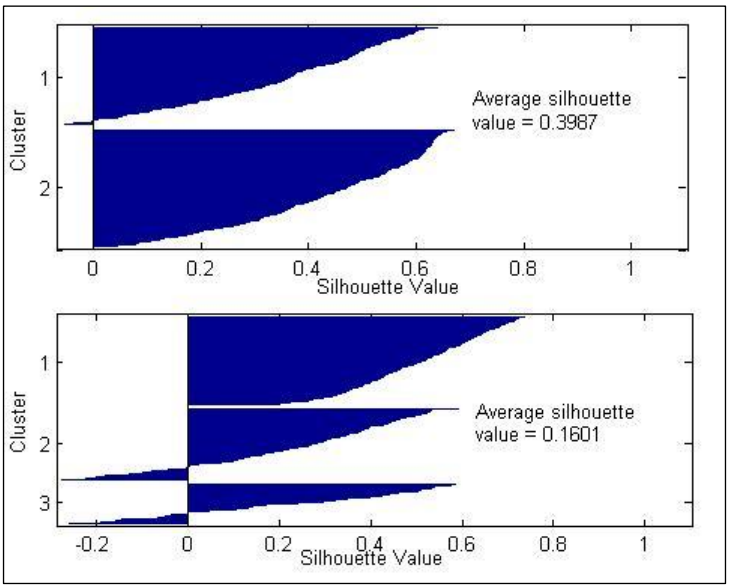


**Figure 3.5 Silhouette value plots for 2 clusters and 3 clusters**

This research determines the number of clusters using silhouette value plots and average silhouette values, as this procedure is free from any assumptions about the distribution of the data. The number of clusters is varied from 2 to 10 and the corresponding average silhouette value is calculated. The number of clusters corresponding to the highest average silhouette value is the chosen as the optimal number of clusters.

Denote the number of TAF clusters chosen using the above procedure as $L^*$. The clusters are represented as $C_1, C_2, .., C_{L^*}$ and contain $|C_1|, |C_2|, .., |C_{L^*}|$ days respectively.

Next, we determine a set of representative capacity scenarios from the arrival capacity profile for the days that have similar TAFs. We performed K-means clustering on the arrival capacity profiles for the days belonging to $C_1, C_2, .., C_{L^*}$. This operation identifies sets of days belonging to the same TAF cluster having similar arrival capacity profiles. Using the same procedure based on average silhouette value and the silhouette value plots used to determine the number of TAF clusters, we choose the number of arrival capacity profile clusters within $C_1, C_2, .., C_{L^*}$.

Denote the number of clusters having similar arrival capacity profiles within $C_k$ as $j_k$ $(k \in \{1,2,..,L^*\})$. Therefore, the $k^{th}$ cluster $C_k$ contains $C_{k,1}, C_{k,2}, .., C_{k,j_k}$ arrival capacity profile clusters, each having $|C_{k,1}|, |C_{k,2}|, .., |C_{k,j_k}|$ days $(|C_{k,1}| + |C_{k,2}| + .. + |C_{k,j_k}| = |C_k|)$. Denote $A^i_{k,m}$ as the arrival capacity profile of the $i^{th}$ day within cluster $C_{k,m}$ $(k \in \{1,2,..L^*\}; m \in \{1,2,..j_k\})$

The set of capacity scenarios associated with $C_k$ are the averages of the arrival capacity profiles of the days within clusters $C_{k,1}, C_{k,2}, .., C_{k,j_k}$. Therefore $C_k$ is associated with $j_k$ number of probabilistic capacity scenarios. The probability of the scenarios is proportional to the number of days within $C_{k,1}, C_{k,2}, .., C_{k,j_k}$ respectively.

The scenarios and their probability within cluster $C_k$ are determined by the equation (3.7). In equation (3.7) the square brackets indicate the nearest integer operation and $\mathbb{I}_{C_k}$ is an event implying that the day-of-operation in classified in cluster $C_k$

$$S_{k,m} = \left[ \sum_{i \in C_{k,m}} A^i_{k,m} \Big/ |C_{k,m}| \right] \qquad P(S_{k,m} | \mathbb{I}_{C_k}) = |C_{k,m}| \Big/ |C_k| \qquad 3.7$$

$$k \in \{1,2,..,L^*\}; m \in \{1,2,..j_k\}$$

We employ these clustering results to generate day-of-operation capacity scenarios from a TAF as follows. First, the TAF is decomposed into a quarter-hour series and standardized with the historical mean and standard deviation. The day is then assigned to one of the TAF clusters $C_k$ $(k \in \{1,2,..,L^*\})$ based on the similarity of its TAF to the cluster centroids. Based on the classification, the capacity scenarios for the day-of-operation would have capacity scenarios determined by equation (3.7).

### 3.3.3 Fog Burn-off Time Clustering

This method of scenario generation requires the STRATUS forecast that is issued exclusively for SFO, and is therefore applicable only for SFO.

We construct the scenarios based on the analysis of the STRATUS forecast generated at 8:00am PDT for 183 days in the summer months of 2004 to 2006. We choose the 8:00am forecast because it is the first of the day for which predictions from the satellite statistical model become available, and is more reliable that the earlier ones.

We concentrated on the days when the actual fog burn-off time, $\tau$, was between 9:30am and 11:30am PDT as the number of days outside this time interval were few. We divided this interval, into eight-quarter-hour intervals, where each interval, $k$ $(k \in 1,2,..8)$, is defined by a lower time boundary, $t_{k,min}$, and an upper time boundary, $t_{k,max}$. The historical days that have

the actual fog burn-off in the same interval are grouped together. Thus we have eight clusters of days. A cluster $C_k$ contains the groups of days that experienced burn-off in time interval k . In other words a cluster, $C_k$ -contains the days when the actual fog burn-off time was between $t_{k,min}$ and $t_{k,max}$ i.e. $t_{k,min} < \tau \leq t_{k,max}$ .

The clusters, the numbers of days in each, $|C_k|$, and the associated fog burn-off time intervals, $t_{k,min}$ and $t_{k,max}$ , are given in table 3.2.

**Table 3.2 Clusters, number of days and time boundaries according to fog burn-off time interval**

| Interval ($\boldsymbol{k}$) | Cluster ($\boldsymbol{C_k}$) | Lower time boundary ($\boldsymbol{t_{k,min}}$) | Upper time boundary ($\boldsymbol{t_{k,max}}$) | Number of days ($\lvert\boldsymbol{C_k}\rvert$) |
|---|---|---|---|---|
| 1 | $C_1$ | 9:30am | 9:45am | 15 |
| 2 | $C_2$ | 9:45am | 10:00am | 16 |
| 3 | $C_3$ | 10:00am | 10:15am | 16 |
| 4 | $C_4$ | 10:15am | 10:30am | 11 |
| 5 | $C_5$ | 10:30am | 10:45am | 24 |
| 6 | $C_6$ | 10:45am | 11:00am | 22 |
| 7 | $C_7$ | 11:00am | 11:15am | 18 |
| 8 | $C_8$ | 11:15am | 11:30am | 15 |

Let $A_k^i$ ($k \in \{1,2,..,8\}$) be the arrival capacity profile of the $i^{th}$ day in $C_k$ . We determine the capacity scenario, $S_k$, for $C_k$ as the average of the arrival capacity profiles of the days within that cluster $C_k$ , rounded to the nearest integer. In equation (3.8), the square brackets indicate the nearest integer operation.

$$S_k = \left[\frac{\sum_{i \in C_k} A_k^i}{|C_k|}\right];\ k \in \{1,2,..8\} \qquad 3.8$$

Figure 3.6 shows the scenarios obtained from this averaging.

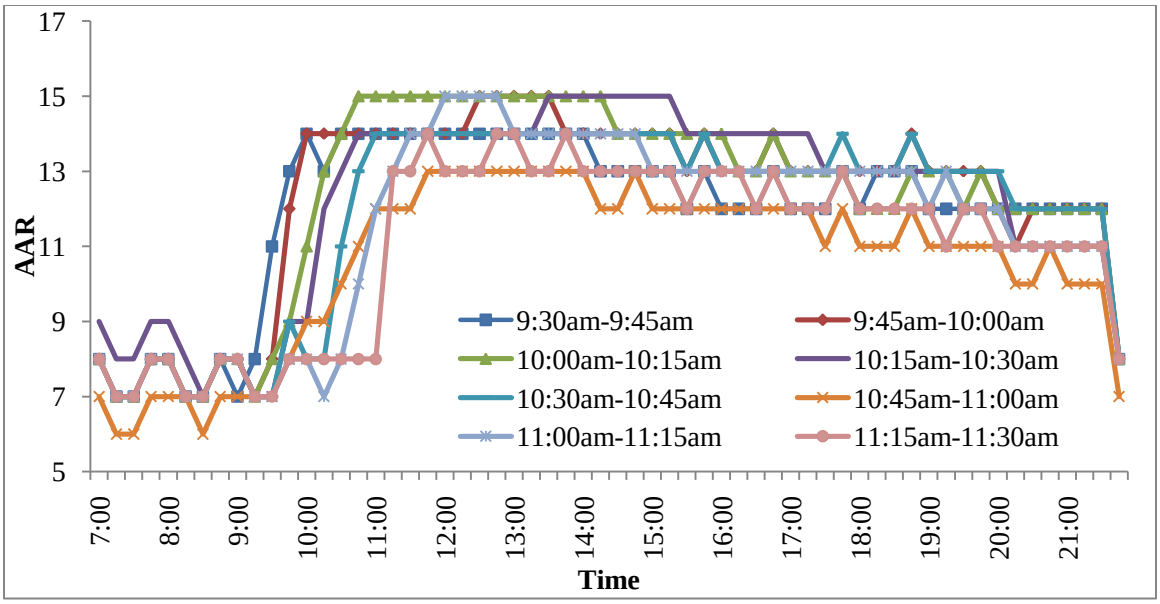


Figure 3.6 Capacity scenarios from Fog Burn-off time Clustering for SFO

Figure 3.7 focuses on the capacity scenarios in periods around burn-off. It shows that the AAR does not increase abruptly after burn-off, but rather increases gradually over a transition period lasting approximately 45 minutes. Discussion with air traffic managers reveals that, at the onset of improving conditions, they are unsure of the exact stratus burn-off time and therefore are reluctant to immediately call the full VMC arrival rate. If they did so, and the stratus temporarily returned, this would result in an excess of aircraft unable to land, compromising safety. The gradual increase in the called arrival rates around the burn-off time thus reflects the risk mitigation behavior of the air traffic managers.

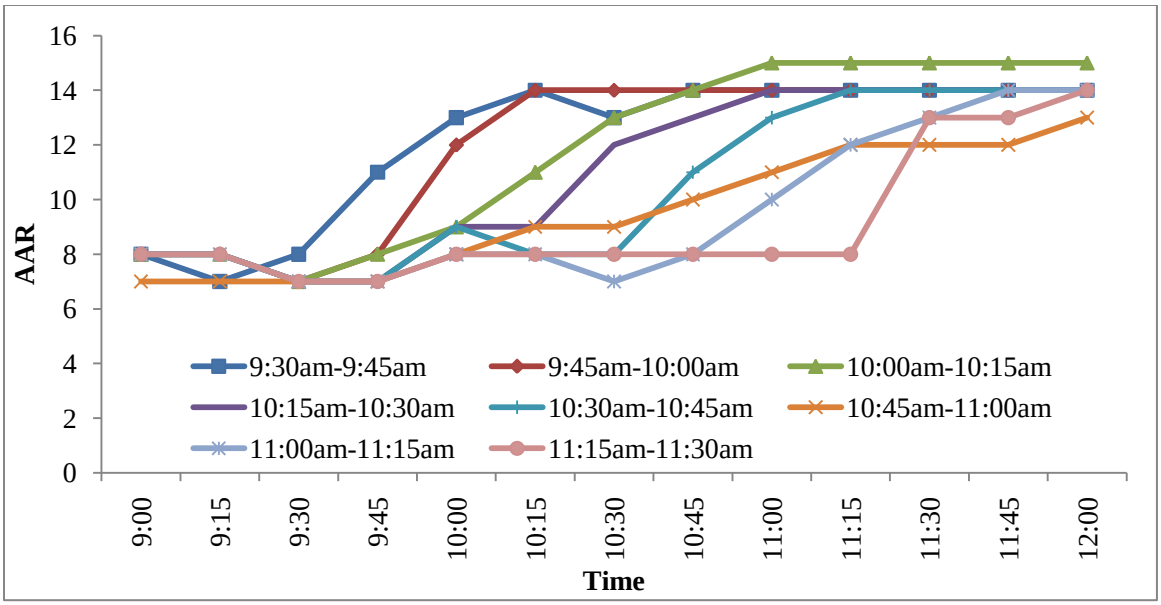


**Figure 3.7 Capacity scenarios from Fog Burn-off time clustering for SFO for period around fog burn-off**

As mentioned in Section 3.2.3 the output from STRATUS also includes the probability of the fog burning off before 10am, 11am, 12pm and 1pm. This probabilistic forecast is essentially the value of the cumulative distribution function (CDF) for the fog burn-off time at 10am, 11am, 12pm and 1pm, conditioned on the forecast burn-off time, as estimated from historical data. We use this probabilistic forecast to determine a-priori probability that a given day belongs to each cluster.

We linearly interpolate the STRATUS CDF between 0:00am, 10am, 11am and 12pm to obtain the CDF values for burn-off for any quarter-hour period between 9:30am and 11:30am PDT. We assume that at 0:00am the value of the CDF is 0. We then estimate the CDF, $F_\tau(t)$, for the $t$ values at 9:30am, 9:45am,.., 11:30am by linearly interpolating the probabilities provided in the STRATUS forecast. The "raw" probability of the fog burn-off time for given day is in the time interval associated with cluster $C_k$ , $P_k^*$, is determined by equation (3.9). As an illustration, figure 3.8 shows the linearly interpolated CDF for 6-17-2004 constructed using probabilities mentioned in table 3.1.

From the CDF, the probability of burn-off in a quarter-hour cluster,$C_k$ , is determined by equation (3.9):

$$P_k^* = P\left(t_{k,min} < \tau \leq t_{k,max}\right) = F_\tau\left(t_{k,max}\right) - F_\tau\left(t_{k,min}\right);\ k \in \{1,2,..8\} \qquad 3.9$$

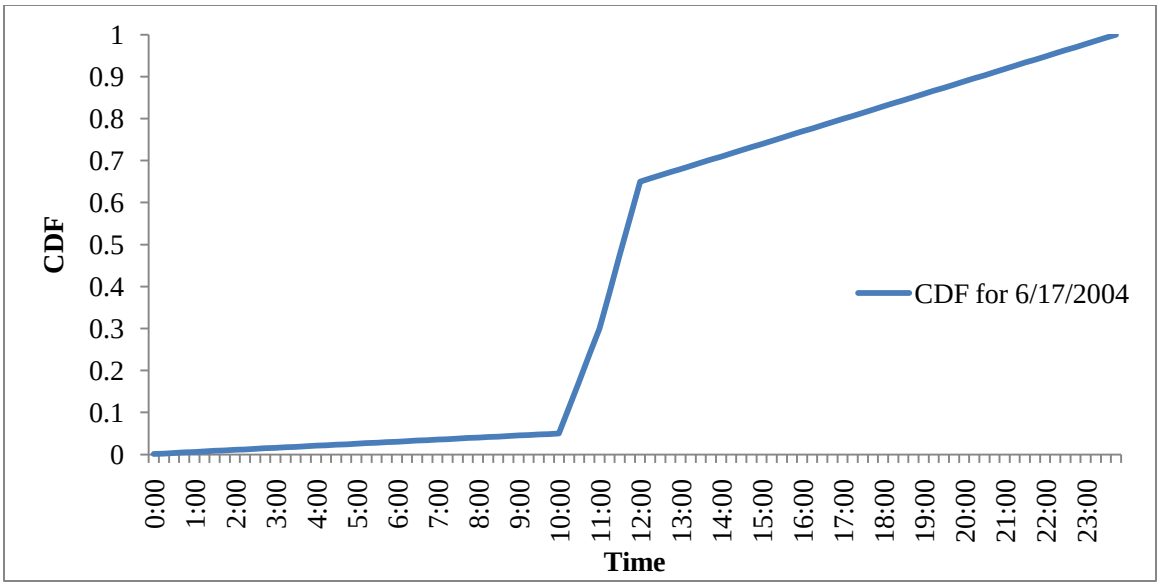


Figure 3.8 Linearly interpolated CDF for 6-17-2004

The probabilities of burn-off in a quarter-hour cluster for 6-17-2004 is given in table 3.3.

Table 3.3 Probability for burn-off in a quarter-hour period for 6-17-2004

| Time period ($k$) | "Raw" probability ($P_k^*$) |
|---|---|
| 1 | 0.00122 |
| 2 | 0.00122 |
| 3 | 0.0625 |
| 4 | 0.0625 |
| 5 | 0.0625 |
| 6 | 0.0625 |
| 7 | 0.0875 |
| 8 | 0.0875 |

We now renormalize these probabilities so that they sum to 1, in effect ignoring the rare cases when the burn-off time is before 9:30am or after 11:30am. We assign the resulting probabilities to the capacity scenarios calculated from the historical days.

$$P(S_k) = \frac{P_k^*}{\sum_{i=1}^{8} P_i^*} \; ; \; k \in \{1,2,..8\} \qquad 3.10$$

Table 3.4 shows the renormalized probability of scenarios for 6-17-2004 that are obtained after equation (3.10)

Table 3.4 Probability for scenarios for 6/17/2004

| Time period ($k$) | Probability $P(S_k)$ |
|---|---|
| 1 | 0.0028 |
| 2 | 0.0028 |
| 3 | 0.146 |
| 4 | 0.146 |
| 5 | 0.146 |
| 6 | 0.146 |
| 7 | 0.204 |
| 8 | 0.204 |

In summary, we have generated eight probabilistic capacity scenarios corresponding to the eight quarter-hour clusters as shown in Figure 3.6. The scenarios are determined from the realized arrival capacity profiles of the days that have fog burn-off times in a given quarter-hour interval. The probabilities of the scenarios are obtained by linearly interpolating the fog burn-off probabilities from the STRATUS forecast of the day-of-operation. Thus, we can determine the probabilities of the scenarios by linearly interpolating the burn-off probabilities included in the day-of-operation STRATUS forecast.

## 3.4 Reference cases

This section describes the methodology for generating the probabilistic capacity scenarios for the two reference cases. In the first case, the scenarios are generated without any weather forecast information while in the second case we assume that air traffic managers have perfect foresight about the arrival capacity profile.

### 3.4.1 Naïve Clustering

This method of scenario generation can be applied to any airport and it does not require any weather forecast information. Accordingly, this reference case is used as a base case to quantify the benefits of using forecasts in decision making and is similar to that described in reference [**34**]. The methodology generates probabilistic capacity scenarios from realized historical capacity without any weather forecast information. A scenario is the average of a group of similar arrival capacity profiles.

Groups of days having similar capacity profiles are determined using K-means clustering. The days within the same cluster have similar arrival capacity profiles. The similarity between two profiles is gauged by the Euclidean norm of the difference between the arrival capacity profiles. A smaller Euclidean norm indicates a greater similarity.

K-means clustering splits the set of day based on the arrival capacity profile in a predefined number of clusters, $L$, where each cluster $C_k$ contains $|C_k|$ days ($k \in 1,2,..L$). $A_k^i$ is the capacity profile of the $i^{th}$ day in the $k^{th}$ cluster. The average of the arrival capacity profiles of the days in a cluster represents the capacity scenario for that cluster and is given by equation (3.11). In this equation the square brackets indicate the nearest integer operator.

$$S_k = \left[ \sum_{i \in C_k} A_k^i \Big/ |C_k| \right] ; \; k \in \{1,2,..L\} \tag{3.11}$$

The probability of each scenario is given by the fraction of the days in each cluster.

$$P(S_k) = \frac{|C_k|}{D} ; \; k \in \{1,2,..L\} \tag{3.12}$$

We observed that K-means clustering on the arrival capacity profiles is susceptible to local minima resulting in clusters with poor silhouette values. To avoid the local minima, we proceeded as follows. First, we specified a number of clusters. Next we chose 100 sets of random starting points in $\mathbb{R}^{60}$, where each coordinate of each point is drawn from the uniform distribution between the quarter-hour IMC capacity and the quarter-hour VMC capacity for an airport, corresponding to the minimum and maximum AAR values. The number of points in the set corresponds to the number of clusters. We then performed K-means for each set of starting points. We repeated this procedure for each number of clusters between two and 10. From the 900 sets of clusters generated, we chose the set with the highest average silhouette value.
We call this procedure Naïve Clustering as it clusters the scenario profiles without any weather information.

### 3.4.2 Perfect Information (PI)

This reference case provides a lower bound on the delay cost resulting from a given demand profile and capacity profile. We assume that the precise evolution of capacity for the entire day-of-operation is known beforehand. A potential capacity-demand imbalance can be predicted accurately and if a GDP is warranted the planned arrival rate will equal the capacity for a period. With the appropriate planned rate, all the delays can be transferred to the ground and associated delay costs will be held to the minimum. The precise capacity forecast enables the best possible planning and avoids unnecessary delays. Thus, in this reference case, the set of scenarios is a single profile that is in fact the realized arrival capacity profile for any historical day.

## 3.5 Realized costs of delays

This section provides a methodology to determine the realized cost of delay by implementing the optimal strategies in a GDP. The cost of the optimal strategies is measured by realized total cost of delay that is the sum of the ground delay costs and the realized air delay costs.

For each airport, we generated the probabilistic capacity scenarios for 45 randomly chosen days in the summer months (May to September) of 2004-2006 using their forecasts. We pretended each of the 45 historical days to be a new day for which we generated the scenarios and then determined the realized costs of delays retrospectively based on the realized capacity.

For all the methods mentioned in this chapter we used the all the days to generate the scenarios and then calculate realized cost of delay for a subset of these days. Thus, in contrast to a real-world application, a day may be used for purposes of both scenario generation and to assess the different scenario generation methods. We performed several experiments in which we removed a given day and recalculated the scenarios, and found, as expected, that this had very little effect on the scenarios or their probabilities. Thus we consider this assessment to be very similar to what would be obtained if assessment days were kept separate from scenario generation days.

For each historical day we determine two sets of PAARs profiles from the SSGDM corresponding to the two sets of scenarios generated by the two scenario generation methodologies: $R_{TC}$ and $R_{NC}$, where the subscripts denote for scenarios based on the TAF Clustering and Naïve Clustering. For SFO, we also generated a third set of PAAR profile, $R_{FB}$, corresponding to the scenarios generated from Fog Burn-off time Clustering using the STRATUS forecast. The generated PAAR can be used to calculate the realized total costs of delay (TCD). The realized total cost of delay is a sum of the total ground delay costs and the total realized air delay costs. The total ground delay can be obtained directly from the SSGDM whereas the total air delay is determined from a deterministic queuing model between the PAAR and the realized capacity.

Figure 3.9 shows a notational queuing diagram between artificially generated scheduled demands, the PAAR determined from SSGDM and the realized capacity. The PAAR matches the realized capacity until time $\tau$. After $\tau$ the PAAR increases but the realized capacity remains low. The area between the scheduled arrivals and the PAAR is the total ground delay and the area between the PAAR and the realized capacity is the total realized air delay. For the Perfect Information case, all the delays would be ground delays and the total delays would be the area between the scheduled arrival and the realized capacity.

The average of the realized total costs of delay ($\overline{\text{TCD}}$) over the 45 days provides us with a tool to empirically quantify realized costs from using scenarios generated from the various scenario generation methods described above. The metric to gauge the benefit of using weather forecasts is $\overline{\text{TCD}}$. Thus $\overline{\text{TCD}}$ provides a basis for evaluating the different scenario generation methods.

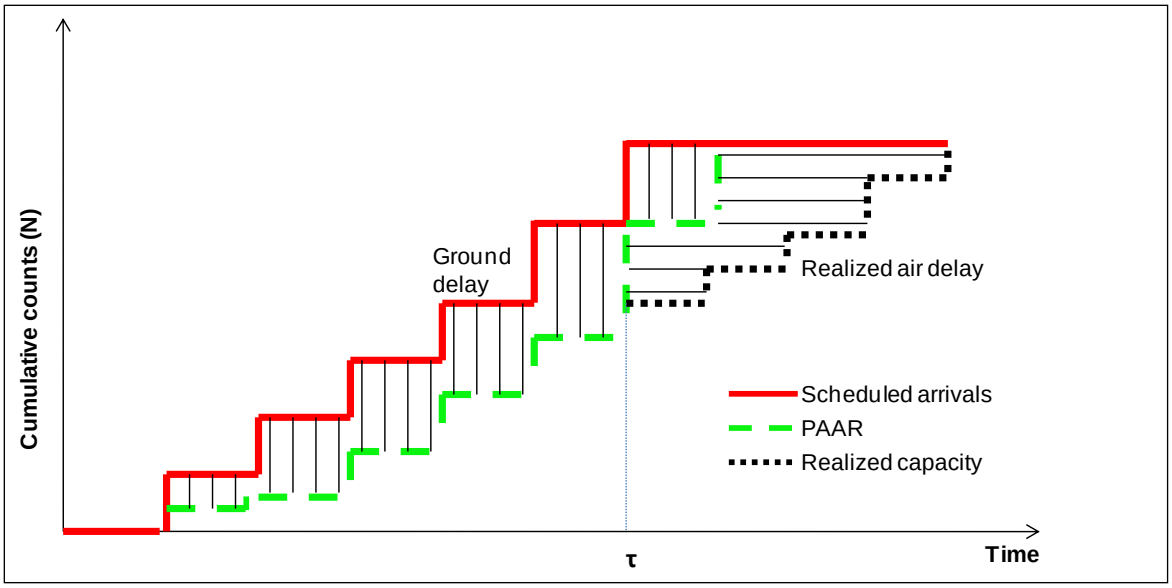


Figure 3.9 Realized capacity, PAAR and scheduled arrivals

# Chapter 4 A design-of-experiment approach for scenario generation

This chapter introduces a new methodology to generate the scenarios from the Terminal Aerodrome Forecast (TAF). It begins by highlighting the limitations of TAF clustering, the scenario generation methodology using the TAFs developed in Chapter 3. It then introduces a new methodology which also uses the TAFs to generate the scenarios. This methodology requires three input tuning parameters which influence the number of scenarios and their probabilities. The optimal values for the parameters are determined using a state-of-the-art optimization algorithm. An outline of this algorithm is also provided in this chapter.

## 4.1 Introduction

The previous chapter develops, TAF Clustering, a methodology using the TAF to generate probabilistic capacity scenarios. The methodology suffers from three limitations:

Firstly, the similarity between the TAFs is determined by comparing the Euclidean norm between the new attributes determined after the PCA operation on the TAF vectors. These new attributes are the linear combination of metrological variables in the TAF vector, prohibiting the direct comparison of metrological attributes in different time periods. This flexibility in comparing the forecasts in proximate periods, called warping, is explained in a later section.

Secondly, the day-of-operation is assigned to the TAF cluster that has the minimum Euclidean distance between its centoid and the TAF for the day-of-operation. This may result in an ambiguous classification if the TAF is equidistant from two or more cluster centroids. The ambiguity in the classification of the day-of-operation might to lead to scenarios that do not adequately represent the uncertainty in the capacity profile. TAF Clustering ignores the individual comparisons between the TAF of the day-of-operation and TAF of historical days.

Thirdly, since the scenarios are generic to a TAF cluster, all days-of-operation classified in the same TAF cluster have the same set of scenarios representing in the uncertainty in the arrival capacity profile. TAF Clustering does not exploit the uniqueness of the TAFs in developing a unique set of scenarios for different day-of-operation.

The above shortcomings associated with TAF Clustering motivate this research in developing a new methodology for scenario generation using the TAF.

## 4.2 Dynamic Time Warping scenarios

This section introduces Dynamic Time Warping (DTW), provides an illustration of DTW, and then discusses the methodology for applying DTW in scenario generation. The section also introduces the three input parameters (Warping Factor, Dimension Factor and Minimum probability threshold) required by this methodology for scenario generation.

As mentioned earlier, it is advantageous to compare the TAF of the day-of-operation with the individual TAFs rather than cluster centroids. There are several techniques that can be used for this comparison including Hidden Markov Models, or measuring the Euclidean distance between TAFs, but this research chooses to perform individual comparisons using Dynamic Time Warping (DTW) due to its ease in implementation.

DTW is an established methodology to study the similarity between two electrical signals, handwriting patterns, or speech patterns. It is particularly useful to match sequences which are translated in time. Recent research has demonstrated that DTW is useful in matching multidimensional time series [**43**]. A multidimensional time series provides data for several attributes, called dimensions, at multiple, discrete time periods. The primary advantage of DTW is it allows features in different, but proximate, time periods to be matched, a procedure referred to as "warping". A TAF can be represented as a multidimensional time series as it contains forecast for seven metrological forecasts attributes for each quarter hour period from 7am to 10pm.

Two multidimensional time series are matched by first constructing a distance matrix. The rows and columns correspond to the number of time periods in the two multidimensional time series. The coordinates $(r, s)$, of any grid cell correspond to the time periods r and s of the first and second multidimensional time series respectively. Each grid cell of the distance matrix contains a cost proportional to the Euclidean norm of the difference between the attributes of the corresponding time periods. The similarity between the two time series is determined by a minimum cost path through the distance matrix from the bottom left corner to the top right corner. The bottom left corner and the top right corner of the grid correspond to the beginning and the end of the multidimensional time series. We constrain the search of the minimum cost path around the diagonal of the distance matrix to match proximate time periods. The summation of the costs in the grid cells on the minimum cost path determines the total cost of the minimum cost path. The total cost is inversely related to the similarity between the two time series. Table 4.1 illustrates the concept for a one-dimensional time series. In table 4.1 below, the numbers in bold are the attributes at consecutive time periods. The italicized numbers in the cells are the costs, i.e. the square of the difference of the attributes between the corresponding time periods. The minimum cost path is highlighted from the bottom left to the top right.

The deviation of the minimum cost path from the diagonal of the matrix occurs when attributes of proximate time periods enough more similar than attributes of the same period. This comparison of attributes in different time periods is referred to as "warping".

**Table 4.1 Dynamic time warping for two artificially generated data, WF=1**

| | Time series 2 → | t=1 | t=2 | t=3 | t=4 | t=5 |
|---|---|---|---|---|---|---|
| Time series 1 ↓ | Feature Vector | **0.69629** | **0.0605** | **0.4851** | **0.72** | **0.4149** |
| t=5 | **0.9709289** | *0.07543* | *0.8289* | *0.236* | *0.063* | *0.3092* |

| | | | | | | |
|---|---|---|---|---|---|---|
| t=4 | **0.1927528** | *0.25355* | *0.0175* | *0.0855* | *0.278* | *0.0493* |
| t=3 | **0.0691176** | *0.39334* | *7E-05* | *0.1731* | *0.4236* | *0.1196* |
| t=2 | **0.3164769** | *0.14426* | *0.0655* | *0.0284* | *0.1628* | *0.0097* |
| t=1 | **0.6997229** | *1.2E-05* | *0.4086* | *0.046* | *0.0004* | *0.0811* |

Although some warping is desired, excess warping is considered undesirable. Excess warping matches attributes of several consecutive time periods of one series with a single time period of the other series. For example, in table 4.1, the attribute of the second and third time period of time series 1, is matched with a single time period of time series 2. Excess warping may lead to counter intuitive alignments [**44**]. We limit warping by multiplying the off diagonal cells by an input parameter referred to as Warping Factor ($WF, \geq 1$). The Warping Factor increases the total cost of the minimum cost path if it deviates from the diagonal. A high value of $WF$ ensures that the minimum cost path has limited deviation from the diagonal resulting in comparisons of attributes closely proximate periods.

The influence on deviation from the shortest path (shown in yellow), when $WF = 10$, for the two artificially generated time series presented earlier in Table 4.1 is demonstrated in Table 4.2. When $WF = 10$ the minimum cost path is aligned on the diagonal.

**Table 4.2 Dynamic time warping for two artificially generated data, WF=10**

| | Time series **2** → | t=1 | t=2 | t=3 | t=4 | t=5 |
|---|---|---|---|---|---|---|
| Time series 1 ↓ | Feature Vector | **0.69629** | **0.0605** | **0.4851** | **0.72** | **0.4149** |
| t=5 | **0.9709289** | *0.7543* | *8.289* | *2.36* | *0.63* | *0.3092* |
| t=4 | **0.1927528** | *2.5355* | *0.175* | *0.855* | *0.278* | *0.493* |
| t=3 | **0.0691176** | *3.9334* | *0.0007* | *0.1731* | *4.236* | *1.196* |
| t=2 | **0.3164769** | *1.4426* | *0.0655* | *0.284* | *1.628* | *0.097* |
| t=1 | **0.6997229** | *1.20E-05* | *4.086* | *0.46* | *0.004* | *0.811* |

For this research, the multidimensional time series being compared are the TAF for the day-of-operation and the TAFs for historical days. The realized capacity profiles of the historic days that have similar TAFs are used as capacity scenarios as determined by the DTW methodology. This research assumes that the probability of a scenario is inversely related to the total cost of the minimum cost path between the historically similar TAF with the TAF of the day-of-operation. Therefore, the greater the total cost, the smaller the probability of the profile and vice versa. The number of capacity profiles is determined by limiting the probability of the least similar profile.

Let $\vec{T}^{O}$ be a multidimensional TAF time series for the day-of-operation of length 60 time periods, where each period is described by the seven attributes. Similarly, let $\vec{T}^{i}$ be the TAF for the $i^{th}$ historical day, also of length 60 time periods with seven attributes for each time period. Let $\vec{T}^{O,r}$ be the seven-attribute TAF forecast for the quarter-hour period $r$ for the day-of-operation. Similarly let $\vec{T}^{i,s}$ be the historical TAF for the quarter hour period s for the $i^{th}$ historical day. Any element $(r, s)$ of the distance matrix is

$$D^{O,i}(r,s) = \begin{cases} \left\|\vec{T}^{O,r} - \vec{T}^{i,s}\right\|^2, r = s \\ \left\|\vec{T}^{O,r} - \vec{T}^{i,s}\right\|^2 \times WF, r \neq s \end{cases} \quad ; \; r,s \in \{1,2,..,60\} \tag{4.1}$$

In equation (4.1), the diagonal grid cells contain the Euclidean distance between the TAF feature vectors and the off diagonal cells contain the Euclidean distance multiplied by the Warping Factor.

The cost of the minimum cost path between $\vec{T}^{O}$ and $\vec{T}^{i}$ is given by $DTW\left(\vec{T}^{O,60}, \vec{T}^{i,60}\right)$ where,

$$DTW\left(\vec{T}^{O,1}, \vec{T}^{i,1}\right) = D^{O,i}(1,1)$$

$$\begin{aligned} &DTW\left(\vec{T}^{O,r}, \vec{T}^{i,s}\right) \\ &= D^{O,i}(r,s) \\ &+ min\left\{DTW\left(\vec{T}^{O,r-1}, \vec{T}^{i,s}\right), DTW\left(\vec{T}^{O,r}, \vec{T}^{i,s-1}\right), DTW\left(\vec{T}^{O,r-1}, \vec{T}^{i,s-1}\right)\right\}; 2 \\ &\leq r,s \leq 60 \end{aligned} \tag{4.2}$$

$$DTW\left(\vec{T}^{O,r}, \vec{T}^{i,s}\right) = \infty, \textit{otherwise}$$

Equation (4.2) is a recursive formula that can be solved with a dynamic program. For the purposes of brevity henceforth, we will write $DTW\left(\vec{T}^{O,60}, \vec{T}^{i,60}\right)$ as $DTW\left(\vec{T}^{O}, \vec{T}^{i}\right)$

This research then ranks the historical days based on the similarity of their TAFs to the TAF for the day-of-operation. $\vec{T}^{<k>}$ is the $k^{th}$ most similar historical TAF to the day-of-operation TAF, so that:

$$DTW\left(\vec{T}^{0},\vec{T}^{<1>}\right) \leq DTW\left(\vec{T}^{0},\vec{T}^{<2>}\right) \leq ..\leq DTW\left(\vec{T}^{0},\vec{T}^{<D>}\right) \quad 4.3$$

The set of probabilistic capacity scenarios is thus the actual realized capacity profiles of the n most similar days. Denoting $A^{<k>}$ as the arrival capacity profile associated with the k$^{th}$ similar day, we define

$$S_k = A^{<k>}; k \in \{1,2,..D\} \quad 4.4$$

As mentioned, the probability of the profile is inversely proportional to the degree of similarity, determined by the total cost of the minimum cost path. We introduce a second input parameter referred to as Dimension Factor ($DF, DF \geq 0$) that determines the importance of the similarity.

$$P(S_k) \propto \frac{1}{DTW\left(\vec{T}^{0},\vec{T}^{<k>}\right)^{DF}} \quad 4.5$$

Equation 4.5 states that the probability of the $k^{th}$ similar profiles is inversely proportional to the total cost raised to $DF$. In our experiments we noticed that $DTW\left(\vec{T}^{0},\vec{T}^{<k>}\right) > 1\ \forall k$, therefore $DF \geq 1$ increases the total cost thereby reducing the similarity between the TAF. A $DF \geq 1$ increases the importance of similarity. Conversely, $DF < 1$, would reduce the total cost increasing similarity. A $DF = 0$ implies that all forecasts are equally similar and therefore, weather forecasts are not useful in controlling delays. Therefore $DF < 1$ decreases the importance of the similarity of the TAF. $DF$ controls the sensitivity of scenario probability to forecast similarity.

We wish to restrict the number of scenarios to $n^*$, by eliminating from consideration those days for which $1\Big/DTW\left(\vec{T}^{0},\vec{T}^{i}\right)^{DF}$ is too small. In other words, we do not consider arrival capacity profiles for the days that have a small degree of similarity. We achieve this through the rule defined by equation (4.6). It determines the number of scenarios by requiring the expression, which will be used below as the basis for calculating the scenario probability, to exceed a threshold, denoted $P_{min}$. $P_{min}$ is the third input parameter. Thus we set:

$$n^* = \max\ k\ ;\ subject\ to\ \frac{\frac{1}{DTW\left(\vec{T}^{0},\vec{T}^{<k>}\right)^{DF}}}{\sum_{i=1}^{k}\frac{1}{DTW\left(\vec{T}^{0},\vec{T}^{<i>}\right)^{DF}}} \geq P_{min} \quad 4.6$$

The scenario probabilities are obtained after normalizing the total cost of the minimum cost path for the $n^*$ days as given by equation (4.7).

$$P(S_k) = \frac{\frac{1}{DTW\left(\vec{T}^0, \vec{T}^{<k>}\right)^{DF}}}{\sum_{i=1}^{k} \frac{1}{DTW\left(\vec{T}^0, \vec{T}^{<i>}\right)^{DF}}} \; ; k \in \{1,2,..n^*\} \tag{4.7}$$

Equation (4.6) determines $n^*$ such that $P(S_{n^*}) \geq P_{min}$ and $P(S_{n^*+1}) \leq P_{min}$

In summary, the scenarios are the realized capacity profiles for the days that have similar TAFs. This methodology requires three input parameters which affect the number and the probability of the scenarios. The effect of the three variables on scenario generation is summarized in the Table 4.3. We call this methodology Dynamic Time Warping scenarios.

**Table 4.3 Summary of the three parameters for scenario generation**

| Parameter | Values | Effect |
|---|---|---|
| Warping Factor (WF) | Low values | Selects days which have similar forecasts for different periods |
| | High values | Selects days which have similar forecasts for similar periods |
| Dimension Factor (DF) | Low values | Decreases sensitivity of scenario probability to forecast similarity |
| | High values | Increases sensitivity of scenario probability to forecast similarity |
| Minimum Probability Threshold ($P_{min}$) | Low values | Selects more scenarios |
| | High values | Selects fewer scenarios |

## 4.3 A design-of-experiment approach to determine parameter values

The three input parameters namely, Warping Factor ($WF$), Dimension Factor ($DF$) and Minimum Probability Threshold ($P_{min}$) are required for scenario generation using DTW Scenarios. The objective of this section is to provide a methodology to determine optimal values for the three parameters that minimize the average realized total costs of delays. A vector of the three parameter values is called a design point and the vector of optimal values is called the optimal design point. We expect the optimal design point to vary across airports. This section discusses why traditional derivative-based optimization methods and heuristic based optimization methods are not employed to determine the optimal design point. It then provides an outline of an algorithm that determines the optimal design point based on Response Surface Methodology (RSM).

In this research we determine the optimal design for each airport based on the sample of the 45 historic days as mentioned in Section 3.5. Ideally, the TAF of a day should be compared with TAFs of more closely preceding days but due to the scarcity of the data we allow comparisons with all the days.

A design point generates a unique set of scenarios for each of the 45 days therefore it influences the average total realized cost of delay. A design point that minimizes the average realized total costs of delays is desired (other objectives may include, minimizing the worst case realized total costs of delays or the median realized total costs of delays). This can be formulated as an optimization problem as shown in equation (4.8). The objective function is to minimize the average total realized costs of delays with the (optional) constraints that the design parameters are between the upper bounds ($UB$) and lower bounds ($LB$).

$$Minimize\ \overline{\mathrm{TCD}}$$

Subject to,

$$LB_{WF} \leq WF \leq UB_{WF} \qquad 4.8$$

$$LB_{DF} \leq DF \leq UB_{DF}$$

$$LB_{P_{min}} \leq P_{min} \leq UB_{P_{min}}$$

There are two challenges associated solving this optimization problem. Firstly, there is no direct formulaic relationship between the design point and the average total realized costs of delay. The average realized total cost of delays is determined after calculating the realized total costs of delays for individual days. The realized total cost of delay for a day depends on its PAAR profile that is obtained by solving the SSGDM with a set of probabilistic capacity scenarios. The set of scenarios are the output from DTW Scenarios. Since, the TAF is unique to a day the scenarios developed from DTW Scenarios are also unique to that day. The capacity profiles that serve as scenarios, the number of scenarios and the probabilities of the scenarios vary across the different days. It is difficult to predict the probability and the number of scenarios directly from the design point prior to implementing the methodology of DTW Scenarios. Therefore, it is difficult to predict the realized total cost of delay and the average for a particular day from the design point.

For this reason, we can imagine the average realized total cost of delay as an output from a black box. The input to the black box is a design point that influences the output. Since, the internal working of the black box uniquely depends on the TAF for a day and is difficult to predict we can pretend that the inside working is unknown to us. The Global Optimization literature refers to the minimization of such objective function in presence of constraints as black box optimizations. For this research, the black box optimization problem is to determine the design point that minimizes the average realize total cost of delays.

Researchers have had limited success in employing traditional derivative based optimization method, that include finite differencing and automatic differentiation, to solve black box optimization problems. The limited success arises as the objective function or the constraints

might be non-smooth leading to unreliable or inaccurate derivatives. In this research it is hard to predict the smoothness of the objective function over the hypercube defined by the constraints.

Derivative-free heuristics based algorithms (genetic algorithms and simulated annealing) are often employed to solve black-box optimizations. These algorithms evaluate the objective function at a large number of design points to determine an optimal design. In our case, however, there is a second challenge: evaluating the objective function at a particular design point is computationally expensive. The average realized total delay costs are determined after three sequential processes: 1) Scenario generation using DTW scenarios. 2) Determining the PAAR profile from the SSGDM using the scenarios as inputs. 3) A deterministic queuing model to determine the realized total delay costs. The scenario generation using DTW scenarios take approximately 60 seconds to generate the scenarios for a day whereas SSGDM and the deterministic queuing models each take less than a second to complete. The relatively large computation time associated with the scenario generation is observed as DTW Scenarios compares a given TAF for a day with all the historical TAFs. This computation time increases with the number of TAFs used in comparisons. For example, the average realized total cost of delay based on 45 days evaluated at a particular design point takes approximately $60 \times 45 = 2700$ seconds or 45 minutes. The repeated sampling of the objective function at different design points is therefore computationally prohibitive making it difficult to employ heuristic based algorithms.

Response Surface Models (RSMs) are a computationally inexpensive, iterative way of minimizing expensive black box optimizations. RSMs approximate the behavior of the underlying expensive black box objective function by evaluating the objective function at a few, initial design points. New candidate design points are intelligently chosen on the response surface where the expensive black-box function is evaluated next. After evaluating the black box function at the new candidate design points the response surface model is updated and the process repeats till a best solution is found or the iterations exceed a threshold value. An optimization algorithm for black-box functions using RSM is given below. This research makes use of the algorithm provided in reference [**45**] to determine the parameter values. The algorithm described in can be decomposed in five major steps :

Step 1: Randomly generate an initial set of $m$ design points $\{x_i\}_{i=1}^{m}$ and evaluate the objective function at these $m$ points. These design points are also called as initial space filling points. In this case each design $x_i$ $(i \in \{1,2,..m\})$ represents a vector of numerical parameter values i.e. $x_i = [WF_i, DF_i, P_{min\ i}]$ . Denote the value of the objective function, the average realized total cost of delay, at these points as $f(x_i)$ ( $i \in \{1,2,..m\}$). Therefore corresponding to each design point an average realized total delay costs is calculated.

Step 2: Fit a response surface model, $s(x)$, using $\{x_i, f(x_i)\}_{i=1}^{m}$. The response surface model, $s(x)$, mimics the influence of the design points on the objective function. The response surface model can be determined using any curve fitting techniques for example, linear regression, polynomial regression, splines or radial basis functions. For the purposes of solving black box optimizations, interpolation basis techniques like kriging or radial basis functions are superior to non-interpolating techniques like linear or polynomial regression [**46**]. In this research we use a radial basis function to develop the response surface.

Step 3: Randomly generate several new candidate design points. Let $\{y_k\}_{k=1}^{l}$ represent a set of $l$ randomly generated candidate design points. Reference [**45**] suggests two ways of generating the random candidate points. The first method generates design points uniformly from the hypercube defined by the upper and the lower bounds of the individual parameters. The second method involves generating random design points that follow a multivariate normal distribution around the current best solution with a predefined covariance structure i.e. the point which currently minimizes the objective function. In this research we follow the second method, generating random points that follow a multivariate normal distribution.

Step 4: From the set of randomly generated points, $\{y_k\}_{k=1}^{l}$, determine a single candidate point to evaluate the objective function for the next iteration. The candidate point is selected based on weighed combination of two competing criterions: Firstly, the candidate point should be far away from the already evaluated points to search for the global best point (distance criterion). Secondly, the candidate point should be selected such that it has a low objective function value on the response surface (response criterion), since the objective is to minimize the objective function. The second criterion competes with the first criterion as the points which minimize the objective function are closer to the current best solution thus promoting a local search.

The values for the two criterions are determined such that they are in between 0 and 1. If the distance criterion value is close to 0 it implies that the candidate point is far from any previously evaluated points thus promoting a global search. If the response criterion is close to 0 it implies that candidate point, on the response surface, has a low objective function value. The weights for the two criterions are selected to be between 0 and 1. The weights can be chosen either in an iterative manner i.e. the weights change for every iteration or they can be fixed for all iterations. In this research we opt for the later and fix the weights for all iterations. Therefore, the weighed combination of a random point is between 0 and 1. The point having the lowest weighed combination is chosen as the candidate point where the objective function would be evaluated next.

Equations (4.9) to (4.18) determine the ideal candidate point where the objective function is evaluated next.

For the $l$ randomly generated points $\{y_k\}_{k=1}^{l}$ determine the objective function value on the response surface, $s(y_k)$ $(k \in \{1,2,..,l\})$. Also compute the maximum and the minimum values on the response surface for the random points.

$$s^{max} = \max_k \; s(y_k)\,;\; k \in \{1,2,..,l\} \tag{4.9}$$

$$s^{min} = \min_k \; s(y_k)\,;\; k \in \{1,2,..,l\} \tag{4.10}$$

The value of the response criterion for a randomly generated point $y_k$ is denoted by $V_k^R$ and given by equation (4.11)

$$V_k^R = \frac{s(y_k) - s^{min}}{s^{max} - s^{min}} \; ; \forall k \in \{1,2,..,l\} \tag{4.11}$$

If at a point, $y_k$, the response surface has a value close to $s^{min}$ then $V_k^R$ has a value closer to 0 and if the response surface has a value close to $s^{max}$ then $V_k^R$ has a value closer to 1.

For each of the $k$ randomly generated points determine the minimum Euclidean distance from each of the previously evaluated points. Defining, $D_k(i)$ is the distance between point $y_k$ and $x_i$

$$D_k(i) = \left|\left|x_i - y_k\right|\right|^2; \; i \in \{1,2,..,n\} \; ; \forall k \in \{1,2,..l\} \tag{4.12}$$

$$\delta_k^{min} = \min_i \; D_k(i) \; ; \; i \in \{1,2,..,n\} \; ; \forall k \in \{1,2,..l\} \tag{4.13}$$

Also determine the following metric given by equation (4.14) and (4.15).

$$\Delta^{min} = \min_k \; \delta_k^{min} \; ; \; \; k \in \{1,2,..,l\} \tag{4.14}$$

And,

$$\Delta^{min} = \max_k \; \delta_k^{min} \; ; \; k \in \{1,2,..,l\} \tag{4.15}$$

Equation (4.14) determines the minimum distance for all random points from all previously evaluated points (min-min). Equation (4.15) determines the largest, minimum distance between all the random points and the previously evaluated points (max-min).

The value of the distance criterion for a randomly generated point $y_k$ denoted by $V_k^D$ is given by equation (4.16).

$$V_k^D = \frac{\Delta^{max} - \delta_k^{min}}{\Delta^{max} - \Delta^{min}} \; ; \forall \; k \in \{1,2,..l\} \tag{4.16}$$

In equation (4.16), if a random point $y_k$ is close to any previously evaluated point then it would have a value close to 1 and similarly if the point is significantly far from any previous point then the value is close to 0.

The weighed sum of the distance criterion and the response criterion for a random point is given by equation (4.17), in which $w^D$ is the weight associated with the distance criterion and $w^R$ is the weight associated with response criterion.

$$W_k = w^D \times V_k^D + w^R \times V_k^R \quad ; \quad \forall k \in \{1,2,..l\}$$
$$w^D + w^R = 1 \; ; 0 \leq w^D, w^R \leq 1 \tag{4.17}$$

The candidate point that becomes the next point for objective function evaluation is a random point in the set $\{y_k\}_{k=1}^{l}$ that has the least weighed sum.

$$x_{n+1} = \min_k \; W_k \tag{4.18}$$

As mentioned earlier there are two ways of choosing the weights: An iterative manner similar to the methodology proposed in reference [**46**] and using fixed weights. The iterative approach contains a vector of values of $w^R$ (and therefore of $w^D$ as $w^D = 1 - w^R$) that start close to 0 and increase progressively to 1. Thus the search starts with a low value of $w^R$ promoting a global search. As new candidate points are evaluated the value of $w^R$ increases and finally terminates at a value of 1 which promotes the local search. Once $w^R$ equals 1, the iterations restart from the beginning of the vector and $w^R$ is again close to 0.

It has been argued that the length of the vector for $w^R$ impacts the optimization process and therefore recent efforts have moved towards fixed weights. When using the fixed weights approach the set of randomly generated candidate points, $\{y_k\}_{k=1}^{l}$, are determined exclusively from a normal distribution around the current best point. The fixed weights are such that $w^R$ is close to 1 ($\approx 0.95$). The low value of $w^R$ promotes local searches around the current best point and the objective function is evaluated in the vicinity of previous best solution. If the objective function value does not reduce by a set threshold the algorithm assumes that a local minimum has been found. The algorithm then restarts and proceeds to generate a new set of random points and a new space of the objective function is explored. Published experimental results indicate that the fixed approach is better and computationally faster for finding global point.

Step 5: Return to step 3, if the current number of iterations or function evaluations are smaller than maximum prescribed. Else report the objective function value and the point where it is achieved.

In this research, we use a radial basis function as a response surface model to mimic the underlying function, average realized total delay costs. The number of initial space filling points is set to 6 points, i.e. $n = 6$ and the number of random points was 6 points per iteration, i.e. $l = 6$. The random points were randomly generated using a normal distribution with mean as the current best solution and the standard deviation as $\sigma_k = 0.1 \times (UB_k - LB_k) \; ; \; k \in WF, DF, P_{min}$ and the correlation was assumed to be 0. The total number of objective function evaluations was set to 200 runs. At the end of the 200 runs the values of WF, DF and $P_{min}$ which have the lowest average realized total cost value was returned. The upper bound and lower bound values for the three parameters are given in Table 4.4 below.

Table 4.4 Upper and lower bounds for parameter values

| Parameter | Lower bound | Upper Bound |
|---|---|---|
| Warping Factor (WF) | 1 | 10 |
| Dimension Factor (DF) | 0.1 | 6 |
| Minimum Probability Threshold ($P_{min}$) | 0.002 | 0.5 |

Ideally, the optimal design should be determined from one set of days (referred as the training set) and the performance of the scenarios should be assessed from a different set of days (referred as the testing set). In this research, due to the lack of number of days, the optimal design and the reported average realized cost of delay are both determined from the same set of days. This may result in over fitting the response surface to average realized total cost of delay.

# Chapter 5 Results, conclusions and future work

## 5.1 Introduction

This chapter provides the conclusions and compares the various scenario generation methodologies. It also provides future research directions. We develop the scenarios and also determine the benefit of using weather forecasts for four US airports: San Francisco International Airport (SFO), Los Angeles International Airport (LAX), Boston Logan International Airport (BOS) and Chicago O'hare International Airport (ORD).

This chapter first presents the probabilistic capacity scenarios for the four airports after implementing the methodology of TAF Clustering and Naïve Clustering. A discussion about the influence of the weather on the scenario profile is also provided. The chapter then provides the numerical values for the three parameters $(WF, DF, P_{min})$ obtained after implementing DTW Scenarios for the four airports. This is followed by a discussion on the parameter values. For each airport, the average realized total costs of delays from the different scenario generation methodologies are also presented. It is shown that scenarios from DTW Scenarios result in the lowest cost of delay amongst all scenario generation techniques. For SFO, we also present the average realized total cost of delay using the STRATUS forecast. We see that the cost of delays using the STRATUS is statistically similar to the cost of delays using TAF and DTW. Finally, certain limitations and future research directions are also discussed.

## 5.2 Probabilistic capacity scenarios from TAF clustering

The probabilistic capacity scenarios generated from TAF Clustering for the four airports are presented in this section. The total number of days $(D)$, the number of TAF clusters $(L^*)$, the number of scenarios in each TAF cluster $(j_k\ ; k \in \{1,2,..,L^*\})$ and the probabilities of the scenarios for each airport is given in tables 5.1 to 5.4. Days that had missing data, either the TAF or the AAR, were removed from the analysis therefore the total number of days vary amongst the airports.

**Table 5.1 Results of TAF Clustering for SFO**

| **SFO** | | |
|---|---|---|
| Number of days $(D)$ | 446 | |
| Number of TAF clusters $(L^*)$ | 2 | |
| Number of scenarios $(j_k)$ | 2 | 3 |
| Probability of scenarios | 0.4604; 0.5396 | 0.2213; 0.4139; 0.3648 |

Table 5.2 Results of TAF Clustering for LAX

| LAX | | |
|---|---|---|
| Number of days $(D)$ | 450 | |
| Number of TAF clusters $(L^*)$ | 2 | |
| Number of scenarios $(j_k)$ | 2 | 2 |
| Probability of scenarios | 0.3631; 0.6369 | 0.2839; 0.7161 |

Table 5.3 Results of TAF Clustering for BOS

| BOS | | |
|---|---|---|
| Number of days $(D)$ | 432 | |
| Number of TAF clusters $(L^*)$ | 2 | |
| Number of scenarios $(j_k)$ | 2 | 2 |
| Probability of scenarios | 0.4454; 0.5545 | 0.6108; 0.3892 |

Table 5.4 Results of TAF Clustering for ORD

| ORD | | | |
|---|---|---|---|
| Number of days $(D)$ | 449 | | |
| Number of TAF clusters $(L^*)$ | 3 | | |
| Number of scenarios $(j_k)$ | 3 | 2 | 3 |
| Probability of scenarios | 0.0879; 0.6428; 0.2692 | 0.1461; 0.8539 | 0.1323; 0.5617; 0.3030 |

The probabilistic capacity scenarios for the four airports are shown in figures 5.1 to 5.9

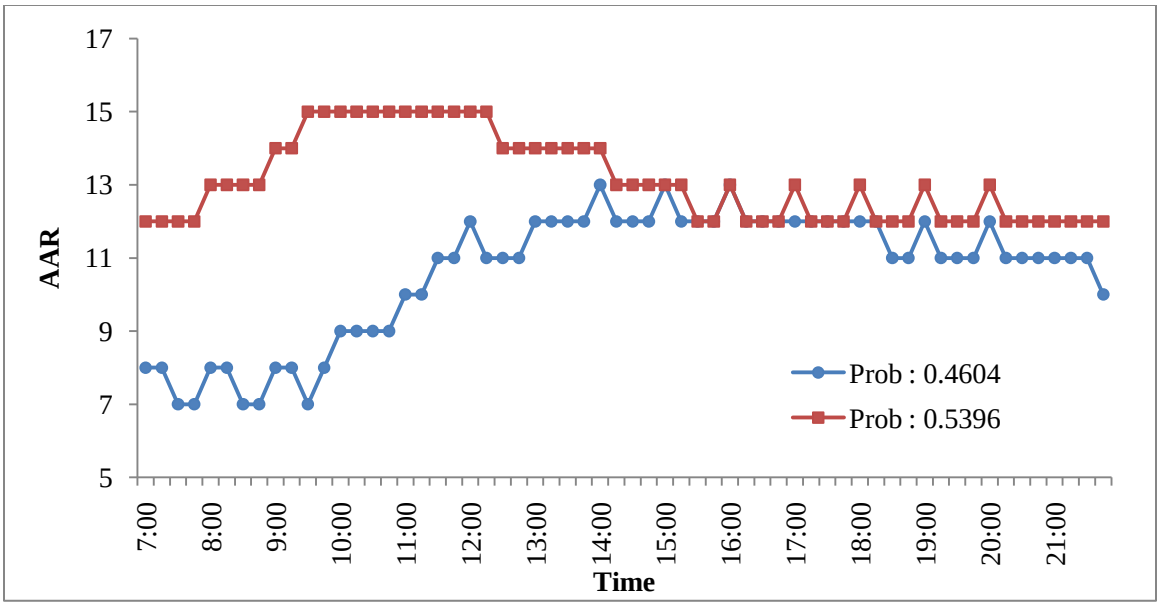


Figure 5.1 Probabilistic capacity scenarios for the days classified in $C_1$ for SFO

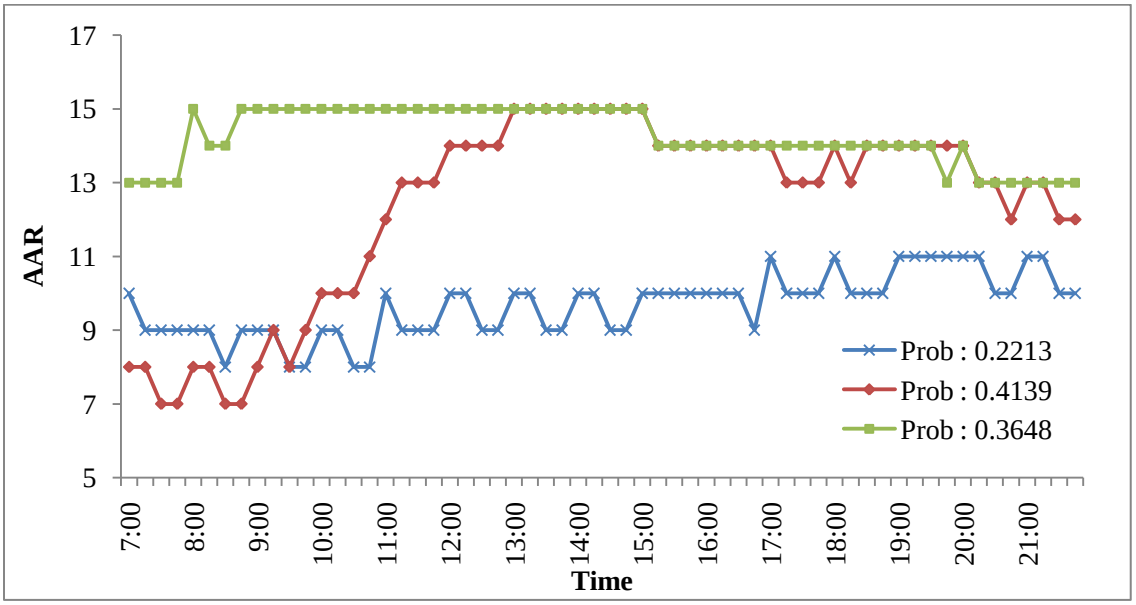


Figure 5.2 Probabilistic capacity scenarios for the days classified in $C_2$ for SFO

In Figure 5.1, the high AAR capacity scenario in red corresponds to the days that have clear weather conditions and stronger, that usual, winds from the San Bruno gap in the afternoon. This prompts a change in runway configuration in the afternoon lowering arrival capacity [**47**]. The gradually increasing AAR capacity scenario in blue corresponds to the days when ceiling and visibility conditions improve as a result of stratus burn-off. Similarly, in Figure 5.2, the high AAR capacity scenario in green corresponds to the days that have clear weather throughout the day. The increasing AAR capacity scenario in red corresponds to the days where ceiling and visibility conditions improve over time. The low AAR capacity scenario in blue corresponds to the days that have poor weather conditions, persisting for the entire day at SFO.

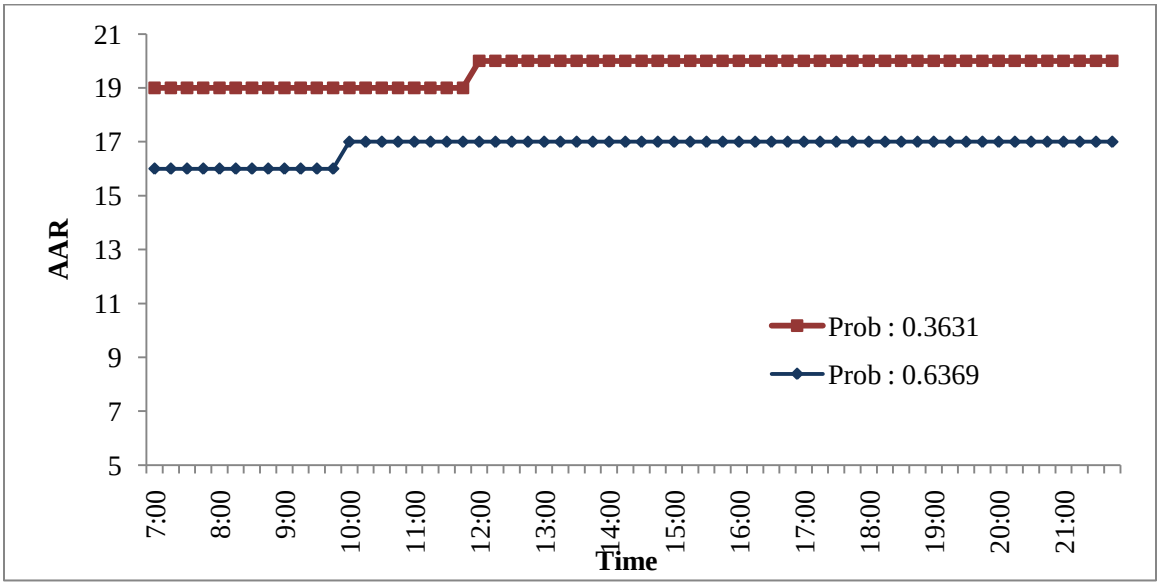


Figure 5.3 Probabilistic capacity scenarios for the days classified in $C_1$ for LAX

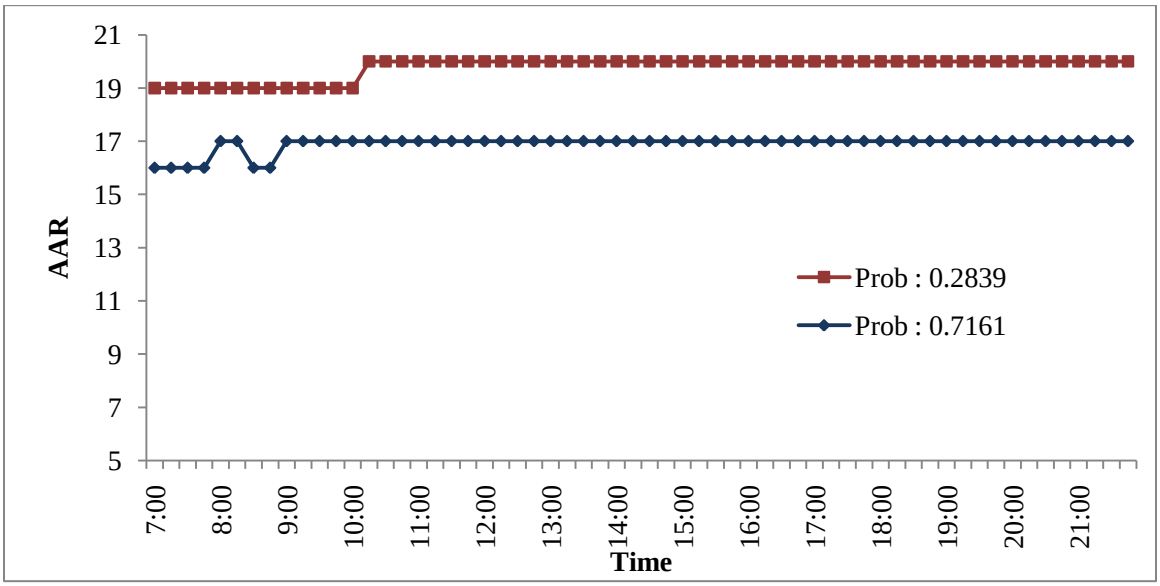


Figure 5.4 Probabilistic capacity scenarios for the days classified in $C_2$ for LAX

In Figures 5.3 and 5.4, the high AAR capacity scenario in red corresponds to the days that have clear weather and LAX is accepting visual arrivals and operating in visual metrological conditions (VMC). The low capacity scenario in blue corresponds to poor weather days when LAX is operating in instrument metrological conditions (IMC). On these days, it is reported that either the cloud ceiling is below 3000ft or the cloud ceilings are present at 8000ft towards the west of the airport [**48**]. The slight increase in capacity for the scenarios in Figure 5.3 occurs later in the day as compared to the scenarios in Figure 5.4. This indicates that that Figure 5.3 represents days where the weather conditions improve later in the day.

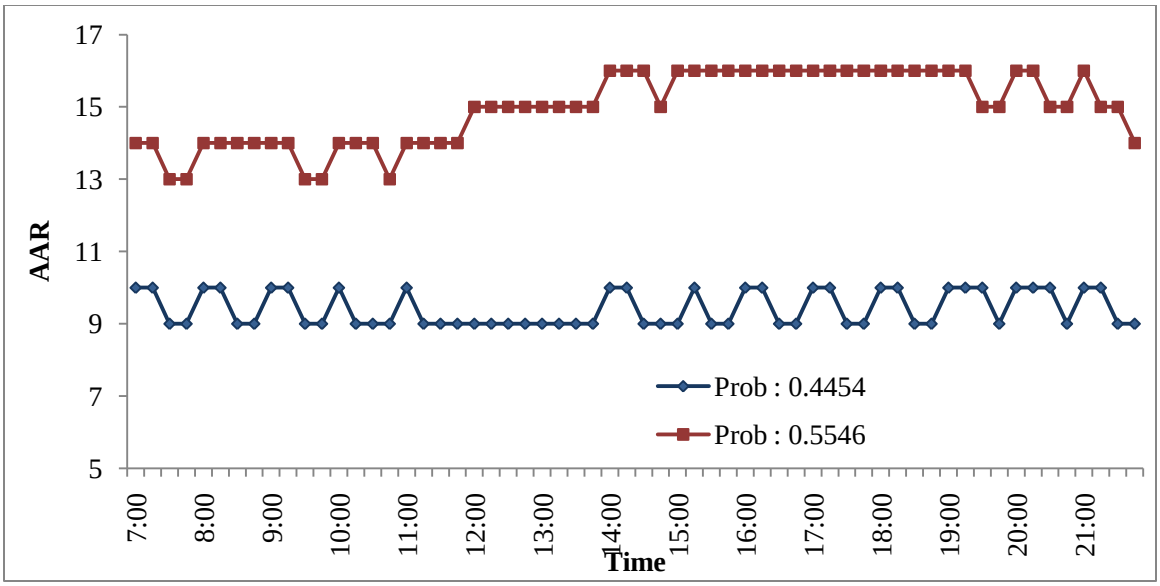


Figure 5.5 Probabilistic capacity scenarios for the days classified in $C_1$ for BOS

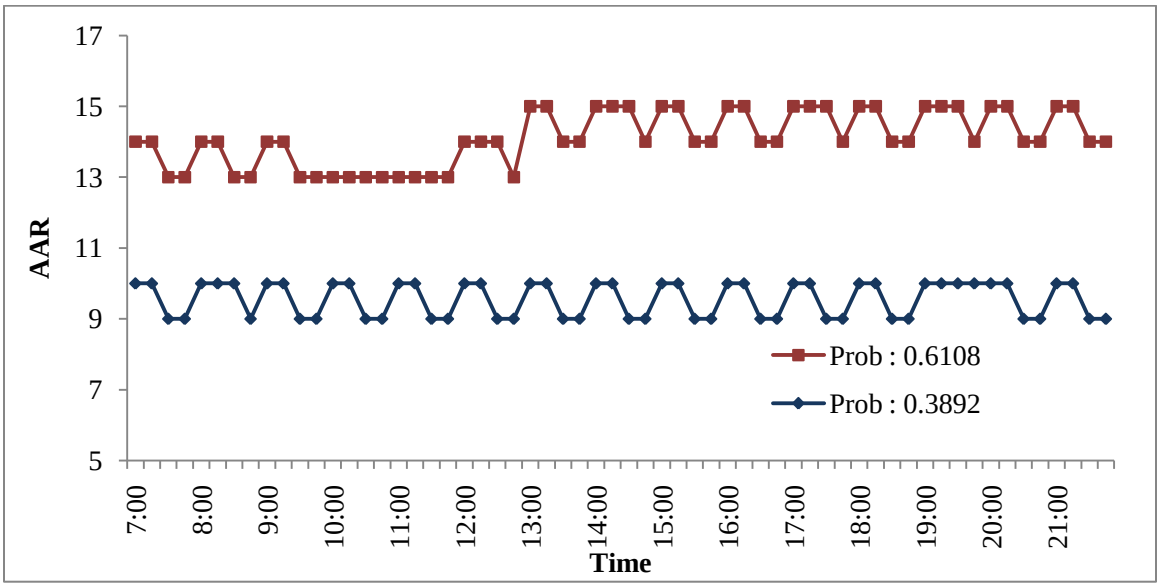


Figure 5.6 Probabilistic capacity scenarios for the days classified in $C_2$ for BOS

In Figure 5.5, the high capacity scenario in red corresponds to days when BOS is operating under VMC. The marginal increase in capacity towards the end correspond to improving ceiling conditions, specifically cloud ceilings become greater than 3500 feet. The increase in cloud ceiling allows a change in the runway configuration thereby permitting higher arrival flows ( [**49**], [**50**]). The lower capacity scenario is blue corresponds to the days when the cloud ceiling is below 3500 feet and BOS is operating under IMC. In Figure 5.6, the high capacity scenario in red corresponds to days that have good weather conditions at BOS and the airport is operating in VMC. The low capacity scenario in blue corresponds to days that have poor weather conditions.

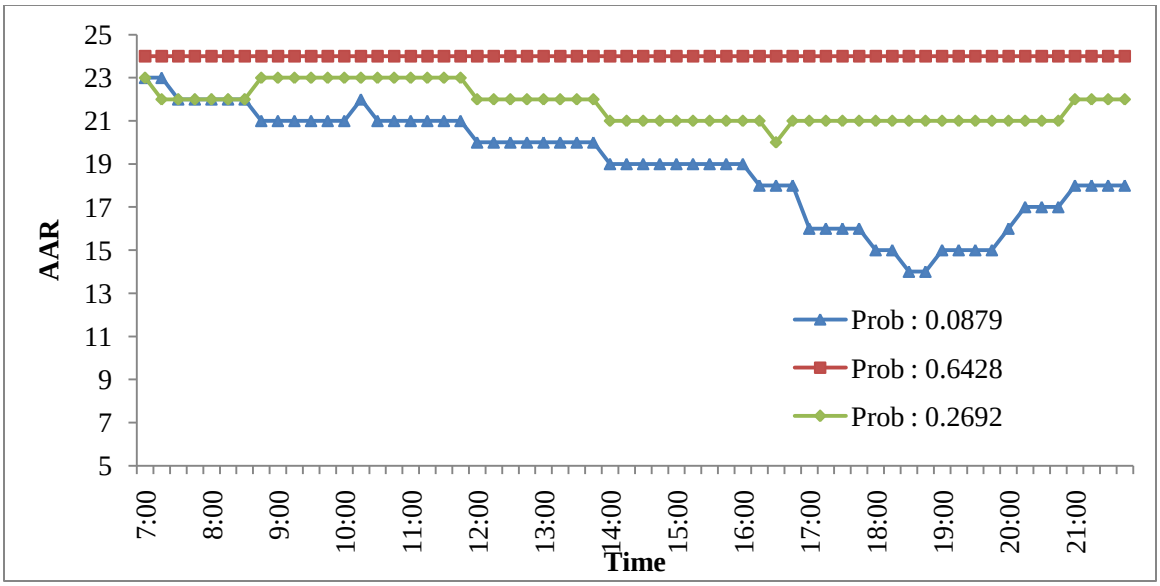


Figure 5.7 Probabilistic capacity scenarios for the days classified in $C_1$ for ORD

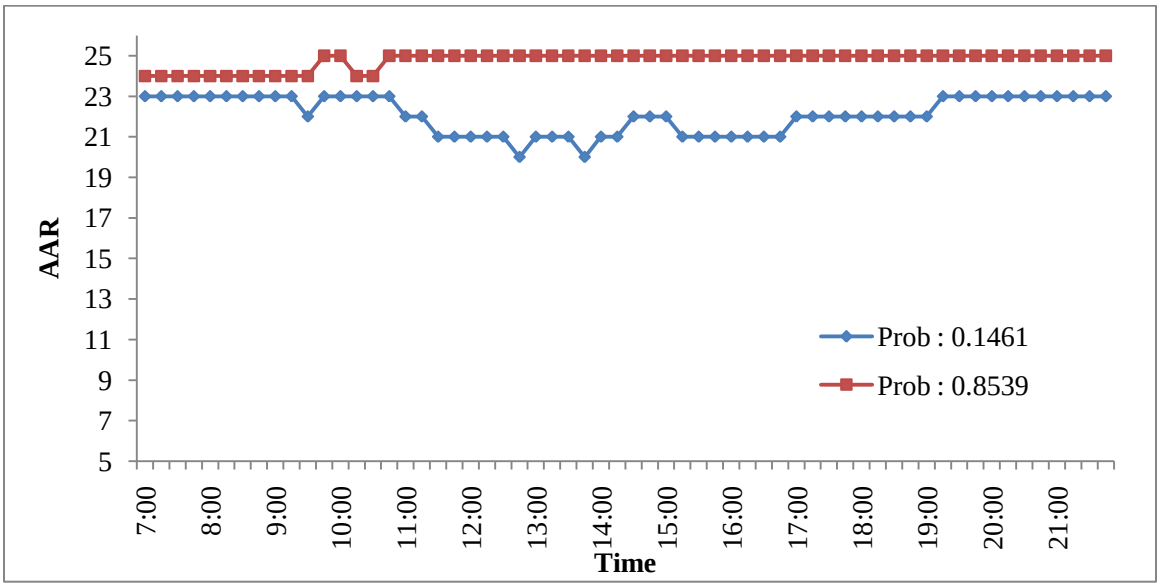


Figure 5.8 Probabilistic capacity scenarios for the days classified in $C_2$ for ORD

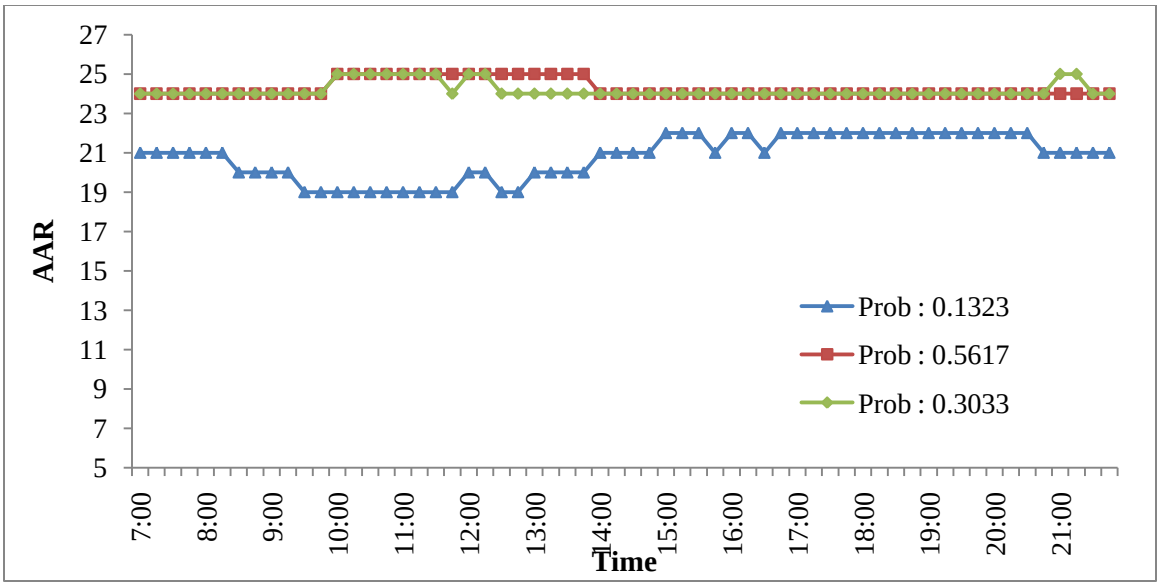


Figure 5.9 Probabilistic capacity scenarios for the days classified in $C_3$ for ORD

In Figure 5.7, the high capacity scenario in red corresponds to day having good weather and ORD is operating in Plan W configuration (the arrivals are on 22R and 27L). This configuration is used on days that enable VMC at the airport and the wind is blowing between 230 to 310 degrees. The green curve corresponds to days when ORD is operating in Plan B configuration. On these days, ORD is operating in VMC while the winds are blowing from 130 to 180 degrees. It is also reported that Plan B is implemented on days when the runways are wet. Finally, the capacity scenario in blue corresponds to days when the weather deteriorates as the day progresses. On these days there are occurrences of afternoon thunderstorms or rain and ORD transitions from VMC to IMC in the afternoon [**51**].

In Figure 5.8, the high capacity scenario in red corresponds to the days that have good weather resulting in the highest arrival rate at ORD. On these days, ORD is operating under Plan X, the preferred arrival configuration (the arrivals are on 27R, 27L and 22R) ( [**49**], [**51**]). The lower capacity scenario in blue corresponds to the days that intermittently have bad weather. ORD operates in VMC at the beginning of the day and transitions into IMC at the onset of bad weather and finally returning to VMC towards the end of the day. This might be attributed to a short duration rain storm. In Figure 5.9, the high capacity scenarios in red and green correspond to days that have good weather but ORD operates in different VMC runway configurations, possibly due to wind direction. The distrin two scenarios require The lower capacity scenario in blue corresponds to days that experience good weather in mid-afternoon. On such days ORD operates in IMC at the beginning of the day and transition into VMC as the day progresses as poor weather conditions improve over time.

## 5.3 Probabilistic capacity scenarios from Naïve Clustering

The probabilistic capacity scenarios for the four airports generated from Naïve Clustering are presented in this section. The number of scenarios ($L^*$)and the probabilities of the scenarios ($P(S_k);\ k \in \{1,2,..,L^*\}$) for the four airports are presented in tables 5.5 to 5.8. The probabilistic capacity scenarios for the four airports are shown in Figures 5.10 to 5.13.

**Table 5.5 Results of Naive Clustering for SFO**

| **SFO** | |
|---|---|
| Number of scenarios ($L^*$) | 3 |
| Probabilities of scenarios $\left(P(S_k)\right)$ | 0.2399; 0.3744; 0.3857 |

**Table 5.6 Results of Naive Clustering for LAX**

| **LAX** | |
|---|---|
| Number of scenarios ($L^*$) | 2 |
| Probabilities of scenarios $\left(P(S_k)\right)$ | 0.6733; 0.3267 |

**Table 5.7 Results of Naive Clustering for BOS**

| **BOS** | |
|---|---|
| Number of scenarios ($L^*$) | 4 |
| Probabilities of scenarios $\left(P(S_k)\right)$ | 0.2013; 0.3680; 0.1088; 0.3217 |

**Table 5.8 Results of Naive Clustering for ORD**

| **ORD** | |
|---|---|
| Number of scenarios ($L^*$) | 2 |
| Probabilities of scenarios $\left(P(S_k)\right)$ | 0.7795; 0.2205 |

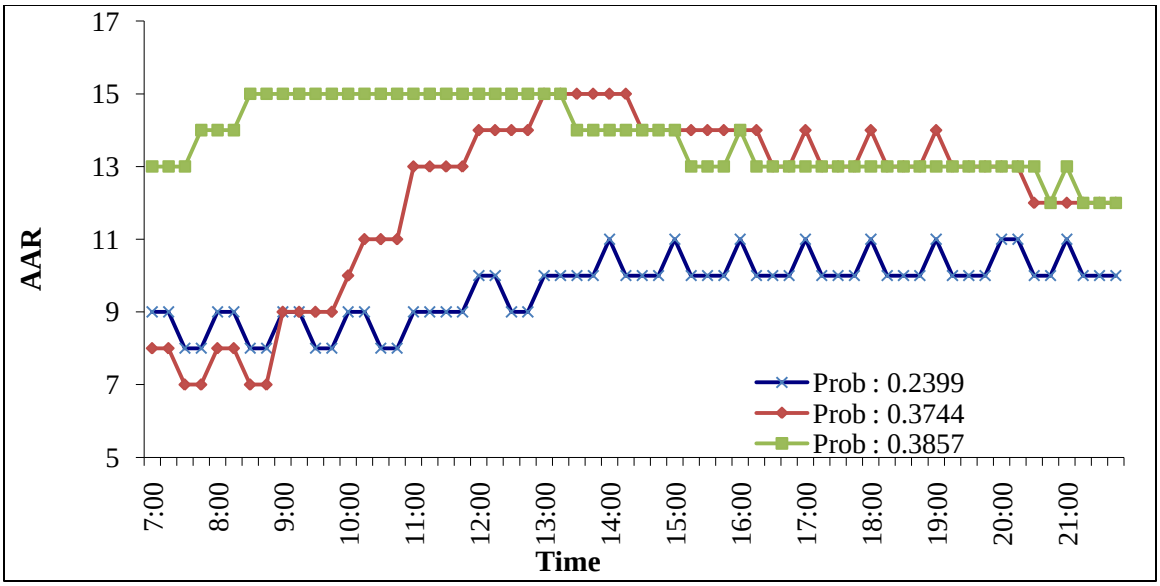


Figure 5.10 Scenarios from Naïve Clustering for SFO

In Figure 5.10, the three scenarios correspond to the three typical types of days observed in the summer at SFO. The first type corresponds to clear weather days that have high AAR and is shown by the green curve. The second type is when the fog sets in the morning and burns off by late morning or early afternoon. This type of day is indicated by the red curve that shows an increase in the AAR in the late morning. The last day type, shown by the blue curve, is when the fog persists and the airport has low landing capacity.

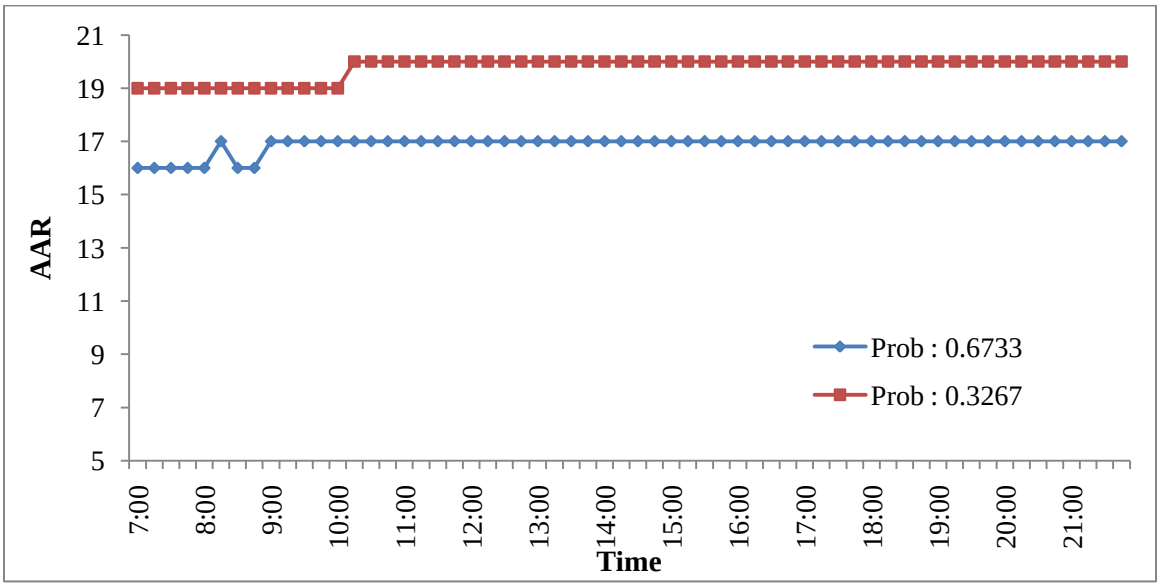


Figure 5.11 Scenarios from Naïve Clustering for LAX

In Figure 5.11, the high capacity scenario in red corresponds to days that have clear weather and LAX operates in VMC. Whereas, the second, lower capacity scenario in blue corresponds to days when LAX is operating in IMC resulting in lower capacity. It is interesting to note that

LAX operates in IMC capacity for almost 67% of the days in the summer and that the capacity is almost consistent throughout the day.

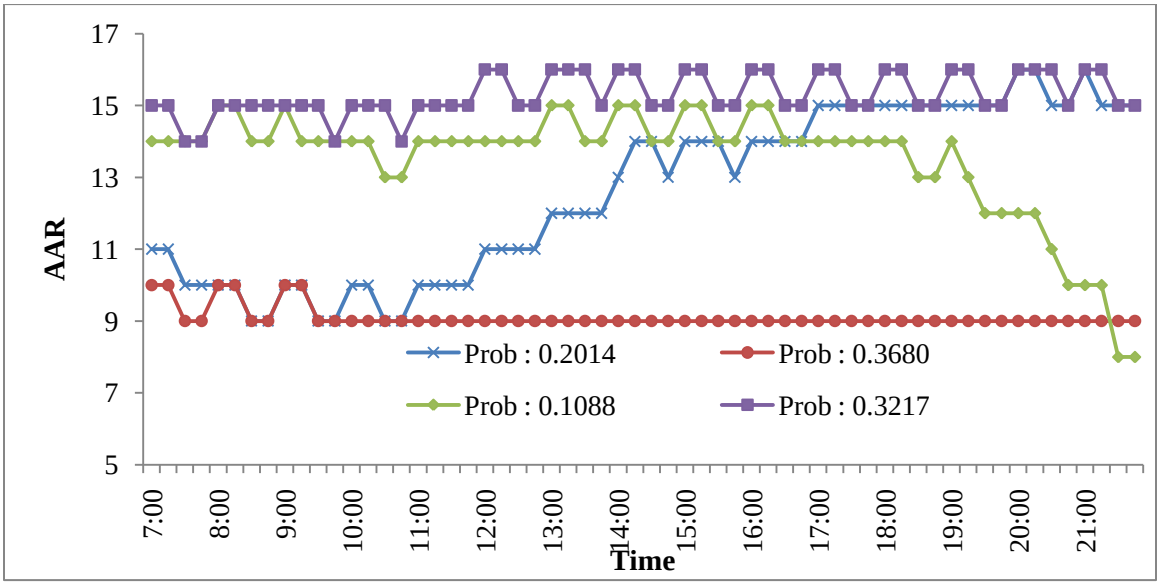


Figure 5.12 Scenarios from Naïve Clustering for BOS

In Figure 5.12, the capacity scenario is purple corresponds to days when BOS is operating under good weather conditions. On such days, BOS operates under VMC and has a high arrival capacity. The capacity scenario in green corresponds to days that experience poor weather in the later part of the day. The operating conditions at BOS transition from VMC to IMC as the day progresses. The capacity scenario in blue corresponds to days that experience improving weather conditions. This may be attributed to improving ceiling and visibility conditions as the day progresses. Finally, the capacity scenario in red corresponds to days that have poor weather conditions throughout the day. On such days BOS operates in IMC for the entire day.

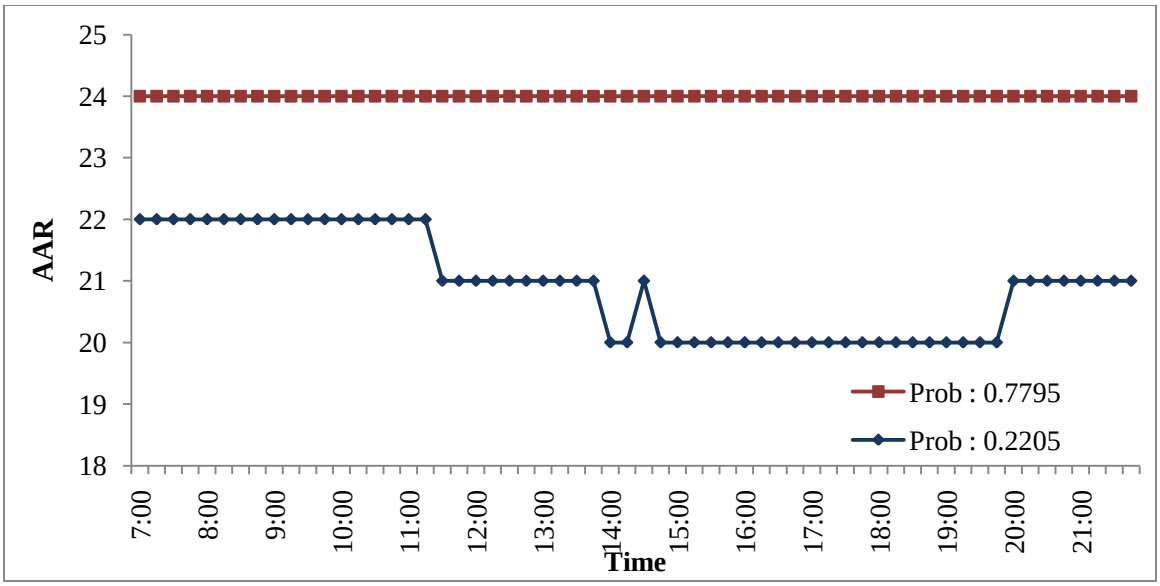


Figure 5.13 Scenarios from Naïve Clustering for ORD

In Figure 5.13, the high capacity scenario in red corresponds to days what have good weather conditions. ORD operates in VMC throughout the day. The capacity scenario in blue corresponds to days that have good weather conditions in the beginning of the day but they worsen in the afternoon. It is hypothesized that afternoon thunderstorms or rain reduce the capacity at ORD. The capacity scenario corresponding to VMC capacity occurs over 75% of days in the summer.

## 5.4 Optimal designs from dynamic time warping scenarios

This section provides the optimal design required for implementing DTW Scenarios described in Section 4.3. A brief discussion on the optimal design is also provided. Once the optimal design is determined, it can be used to generate the probabilistic capacity scenarios for any day-of-operation using the issued TAF and DTW.

Table 5.9 provides the optimal design for the four airports. The range of the number of scenarios observed for the different days at the optimal design across the four airports is also provided.

Table 5.9 Optimal design and number of scenarios for different airports

| Airport | Design Parameters | | | Number of scenarios |
|---|---|---|---|---|
| | $WF$ | $DF$ | $P_{min}$ | |
| SFO | 1.57 | 0.79 | .0023 | 350-400 |
| BOS | 6 | 2.5 | .065 | 12-25 |
| LAX | 1.75 | 2.5 | .035 | 20-25 |

| | | | | |
|---|---|---|---|---|
| ORD | 4.74 | 2.27 | 0.043 | 15-25 |

The optimal design is unique for different airports as each airport has unique operating capacities, demand levels and terminal weather conditions. Therefore the optimal design needs to be determined from historical data before planning arrival operations using DTW Scenarios for any day in the future.

We notice the value of $WF$ varyies significantly across the different airports. $WF$ is the greatest for BOS indicating that there is a high penalty for deviation of the shortest path around the diagonal of the distance matrix. Therefore forecasts of proximate quarter hour periods are compared for scenario generation. The value of $WF$ is minimum for SFO, so that more deviation from this diagonal is tolerated. Forecasts for quarter hour periods further separated in time are compared for scenario generation.

The highest value of $DF$ is for BOS and LAX whereas the lowest value is for SFO. This implies that the scenario probability is more sensitive for TAF similarity for BOS and LAX. For BOS, LAX and ORD, $DF > 1$ implying that similarity of the TAF is important in scenario generation. For SFO, $DF < 1$, decreasing the importance of the similarity of the TAFs. Days that are less similar to the day-of-operation based on the TAF and are included in generating probabilistic capacity scenario.

Finally, the value of $P_{min}$ is, by far, the lowest for SFO, implying that it will have highest number of scenarios. It is surprising to observe that the number of scenarios for the other airports is smaller by an order of magnitude. The precise reason for the high number of scenarios at SFO needs further investigation.

## 5.5 Average realized total cost of delay

This section provides the average realized cost of delay for the different scenario generation techniques for the four airports. As a reminder for each day, the realized cost of delay is calculated by first determining a PAAR using the SSGDM corresponding to each scenario generation methodology. The ground delay is obtained directly from the SSGDM. The realized air delay is obtained by calculating the delay between the realized arrival capacity and the PAAR for that day using a deterministic queuing diagram. We use statistical tests to compare the average realized delay costs from the methodologies. The results allow us to draw some tentative conclusions about the relative performance of the different methods of developing capacity scenarios from weather forecasts that we have developed in this research.

Table 5.10 provides the average realized total cost of delay under the different methodologies of scenario generation: perfect information case, naïve clustering, TAF clustering, DTW Scenarios (abbreviated as $\overline{\text{TCD}}_{\text{PI}}$, $\overline{\text{TCD}}_{\text{Naive}}$, $\overline{\text{TCD}}_{\text{TC}}$ and $\overline{\text{TCD}}_{\text{TDTW}}$). As a reminder, TAF clustering and DTW Scenarios require the TAF to generate the scenarios. For SFO, we also present the average realized total cost of delay using the Fog burn-off methodology (abbreviated as $\overline{\text{TCD}}_{\text{FB}}$). In presenting results here we convert the cost to units of ground delay minutes by multiplying the objective function (equation (2.1)) by 15. The last column shows the benefit of using weather

forecast by calculating the percentage increase in realized delay cost when using the scenarios developed by Naïve clustering over scenarios developed by DTW Scenarios. Column 2 of Table 5.10 shows that the cost under the PI case varies greatly between the different airports, indicating the different capacity and demand imbalances at different airports.

**Table 5.10 Average realized total cost of delay (in ground delay minutes) over 45 days**

| Airport | PI | Naïve | TAF Clustering | TAF-DTW Scenarios | Fog burn-off | $\frac{\overline{TCD}_{Naive} - \overline{TCD}_{TDTW}}{\overline{TCD}_{Naive}} \times 100$ |
|---|---|---|---|---|---|---|
| SFO | 1448 | 3543.50 | 2916.75 | 2708.47 | 2733 | 23.56% |
| LAX | 306.15 | 621.60 | 626.25 | 585.66 | - | 5.78% |
| BOS | 2942 | 9249.62 | 8550.3 | 6557.21 | - | 29.10% |
| ORD | 12755 | 34801 | 33413 | 28996.5 | - | 16.67% |

We performed paired, two tailed t-tests to determine if the differences between the $\overline{TCD}$s are statistically significant. In these tests, the null hypothesis is that the difference between the $\overline{TCD}$s obtained from the different methodologies is zero while the alternative hypothesis is that difference is other than zero. The p-values for the different pairs of $\overline{TCD}$ are given in Tables 5.11-5.14. The values in *italics* indicate the cases in which the null hypothesis is rejected at a significance level of 10%. For SFO we also compare the Fog burn-off methodology with other methods of scenario generation.

**Table 5.11 p-values from the paired t-test for SFO**

| SFO | | | | | |
|---|---|---|---|---|---|
| | **PI** | **Naïve** | **TAF Clustering** | **TAF-DTW Scenarios** | **Fog burn-off** |
| **PI** | - | *3.41E-12* | *6.84E-11* | *1.50E-11* | *2.69E-11* |
| **Naïve** | *3.41E-12* | - | *1.62E-05* | *4.40E-07* | *1.44E-06* |
| **TAF Clustering** | *6.84E-11* | *1.62E-05* | - | *6.98E-05* | *2.79E-04* |
| **TAF-DTW Scenarios** | *1.50E-11* | *4.40E-07* | *6.98E-05* | - | 0.33 |
| **Fog burn-off** | *2.69E-11* | *1.44E-06* | *2.79E-04* | 0.33 | - |

Table 5.12 p-values from the paired t-test for LAX

| LAX | | | | |
|---|---|---|---|---|
| | **PI** | **Naïve** | **TAF Clustering** | **TAF-DTW Scenarios** |
| **PI** | - | *8.61E-04* | *7.23E-04* | *3.18E-03* |
| **Naïve** | *8.61E-04* | - | *0.05* | 0.19 |
| **TAF Clustering** | *7.23E-04* | *0.05* | - | 0.15 |
| **TAF-DTW Scenarios** | *3.18E-03* | 0.19 | 0.15 | - |

Table 5.13 p-values from the paired t-test for BOS

| BOS | | | | |
|---|---|---|---|---|
| | **PI** | **Naïve** | **TAF Clustering** | **TAF-DTW Scenarios** |
| **PI** | - | *7.36E-08* | *4.50E-05* | *1.02E-04* |
| **Naïve** | *7.36E-08* | - | *0.08* | *8.11E-03* |
| **TAF Clustering** | *4.50E-05* | *0.08* | - | *0.06* |
| **TAF-DTW Scenarios** | *1.02E-04* | *8.11E-03* | *0.06* | - |

Table 5.14 p-values from the paired t-test for ORD

| ORD | | | | |
|---|---|---|---|---|
| | **PI** | **Naïve** | **TAF Clustering** | **TAF-DTW Scenarios** |
| **PI** | - | *6.51E-04* | *2.52E-04* | *2.26E-03* |
| **Naïve** | *6.51E-04* | - | 0.42 | *6.47E-03* |
| **TAF Clustering** | *2.52E-04* | 0.42 | - | *0.015* |
| **TAF-** | *2.26E-* | *6.47E-* | *0.015* | - |

| | | | | |
|---|---|---|---|---|
| **DTW Scenarios** | *03* | *03* | | |

Tables 5.10 to 5.14 indicate that probabilistic capacity scenarios derived from DTW Scenarios are better than scenarios devoid of forecast information as $\overline{\text{TCD}}_{\text{TDTW}}$ is statistically lower than $\overline{\text{TCD}}_{\text{Naive}}$ at a significance level of 10%. For all airports except from LAX the above fact is evident from the p-value tables. For LAX, we modify the alternate hypothesis to $\overline{\text{TCD}}_{\text{TDTW}} \leq \overline{\text{TCD}}_{\text{Naive}}$ allowing us to perform a one-tailed paired t-test, the p-value reported in Table 5.12 is reduced by half allowing us to reject the null hypothesis at the significance level of 10%.

When we compare the probabilistic capacity scenarios from TAF Clustering to Naïve Clustering we observe that $\overline{\text{TCD}}_{\text{TC}}$ is lower than $\overline{\text{TCD}}_{\text{Naive}}$ for all airports except LAX. For LAX, $\overline{\text{TCD}}_{\text{TC}}$ is 0.8% higher than $\overline{\text{TCD}}_{\text{Naive}}$ and, furthermore, this difference is statistically different than 0. From this we conclude that TAF Clustering, while probably somewhat better than Naïve clustering, does not generate much value from day-of-operation weather information. This is probably because of the limitations of this approach discussed in Chapter 4.

For SFO, the average realized total cost of delays using the scenarios generated from the STRATUS forecast are marginally higher (0.9%) than $\overline{\text{TCD}}_{\text{TDTW}}$, but the difference is not statistically significant. The use of a specialized weather forecast product—STRATUS—offers no improvement to plan operations. We learned from the developers of the STRATUS that STRATUS is used as an input to generate the TAF for SFO, this may explain the statistically similar costs using the two forecast products.

Our results reveal that the average realized delay costs determined from using TAF and DTW is lower than Naïve clustering for all four airports. This implies that, of the methods we tested, DTW Scenarios is the preferred means of generating probabilistic capacity scenarios incorporating the TAF weather forecast.

In summary, our four case studies suggest that probabilistic capacity scenarios derived from weather forecasts using the DTW method work better for planning operations than scenarios developed without a weather forecast.

We also compare the realized costs from the scenarios based on DTW and Naïve Clustering for individual days. Our hypothesis was that the level of benefit of using weather forecast should vary with the cost under the PI information case. The y-axis of Figures 5.14 to 5.17 represents the difference between the total realized cost of delays using DTW scenarios and Naïve Clustering $(\text{TCD}_{\text{TDTW}} - \text{TCD}_{\text{Naive}})$ for individual days. The individual days are arranged in ascending order of the cost under the perfect information case $(\text{TCD}_{\text{PI}})$ represented on the x-axis. When the difference is positive, it indicates that the scenarios generated using Naïve Clustering perform better that scenarios generated from DTW scenarios. This implies that the scenarios from DTW Scenarios are not useful in controlling the delay costs on that day. Conversely, when the difference is negative, it indicates that scenarios generated from the DTW scenarios outperform the scenarios from Naïve Clustering. From Figure 5.14 for SFO, we observe that the difference is positive for days that have a lower cost under the PI case. The difference becomes negative as the cost under the PI case increases. This suggests that the DTW approach yields the

greatest benefits on days that have a greater capacity demand imbalance. In other words, the TAF reduces the downside risk. The benefit of using the using the TAF on days that have high $TCD_{PI}$ outweighs the loss on days that have low $TCD_{PI}$. Since, however, $TCD_{PI}$ can't be predicted at the beginning of a day, the above observation affirms that the TAF in conjunction with DTW should be used for planning operations on all days-of-operation. If the arrivals are metered using the scenarios from DTW Scenarios, the realized cost would be comparable if arrivals were metered using scenarios from Naïve clustering when $TCD_{PI}$ is low. Other the other hand when $TCD_{PI}$ is high, the metering performed using the scenarios obtained using the DTW methodology would yield a lower cost as compared to Naïve Clustering.

A similar trend is also seen for ORD. For BOS, the trend is less pronounced, but scenarios based on DTW outperform the scenarios from Naïve Clustering for 80% of the days considered in the experiments. There is not clear pattern in the case of LAX perhaps because our sample does not contain any days when conditions at LAX were very poor.

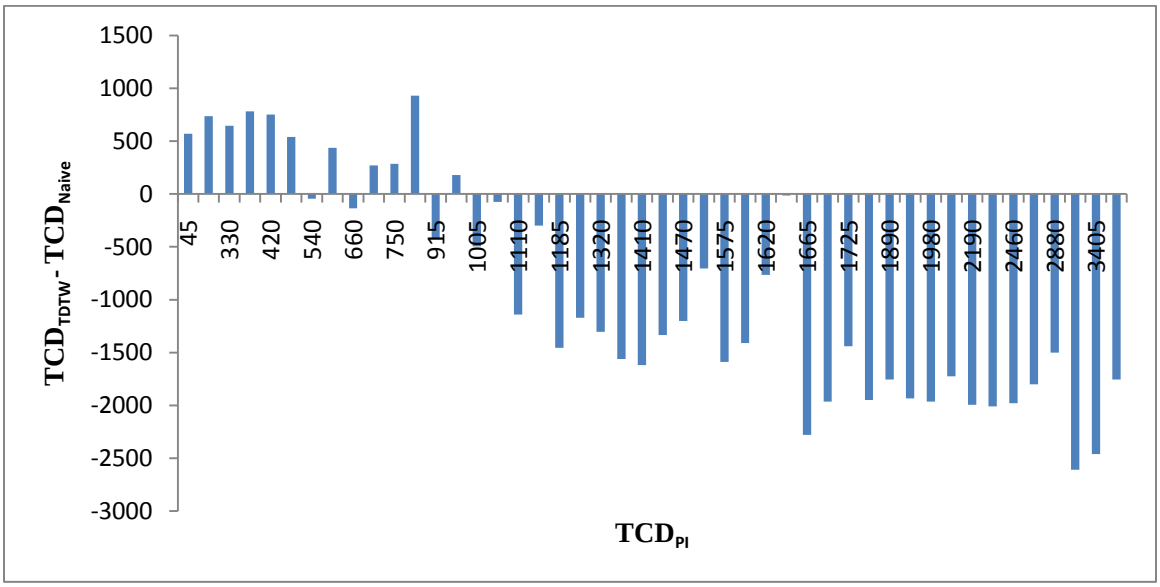


**Figure 5.14 $TCD_{TDTW} - TCD_{Naive}$ versus $TCD_{PI}$ for SFO**

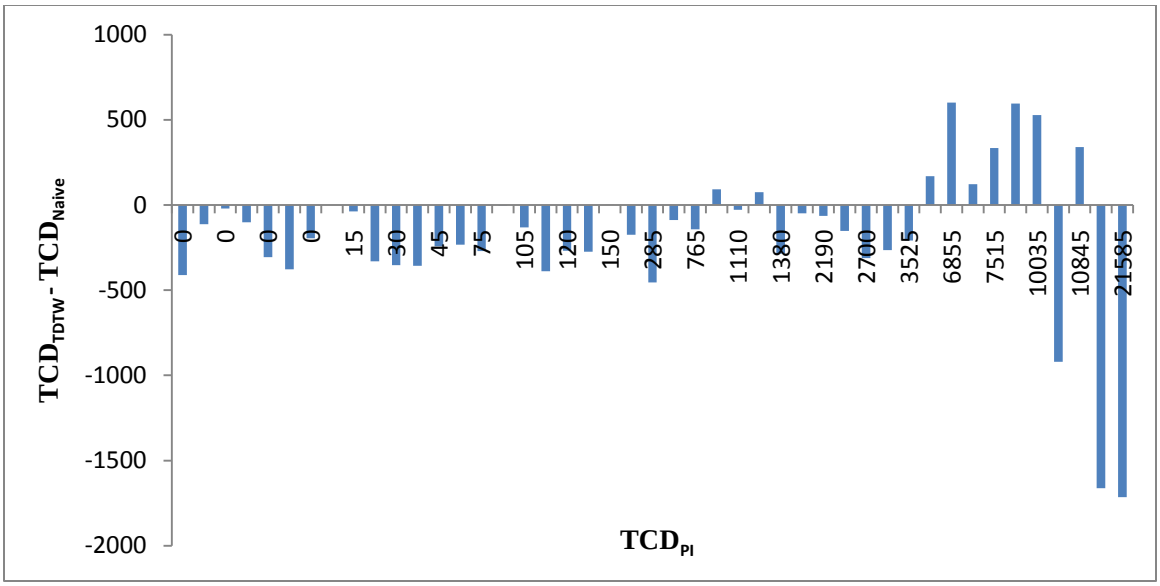


Figure 5.15 $TCD_{TDTW} - TCD_{Naive}$ versus $TCD_{PI}$ for LAX

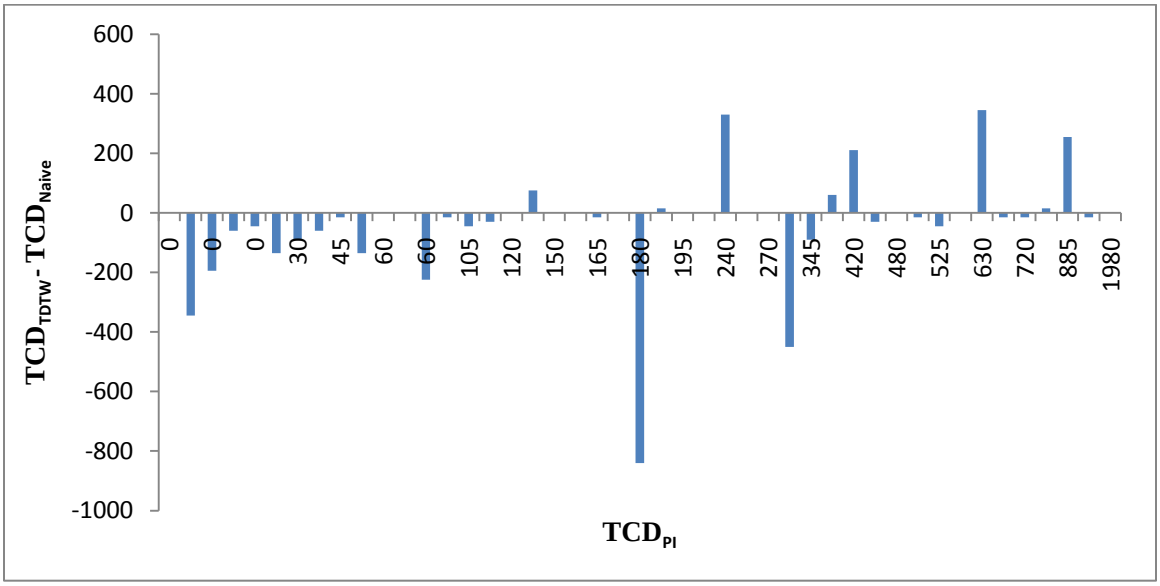


Figure 5.16 $TCD_{TDTW} - TCD_{Naive}$ versus $TCD_{PI}$ for BOS

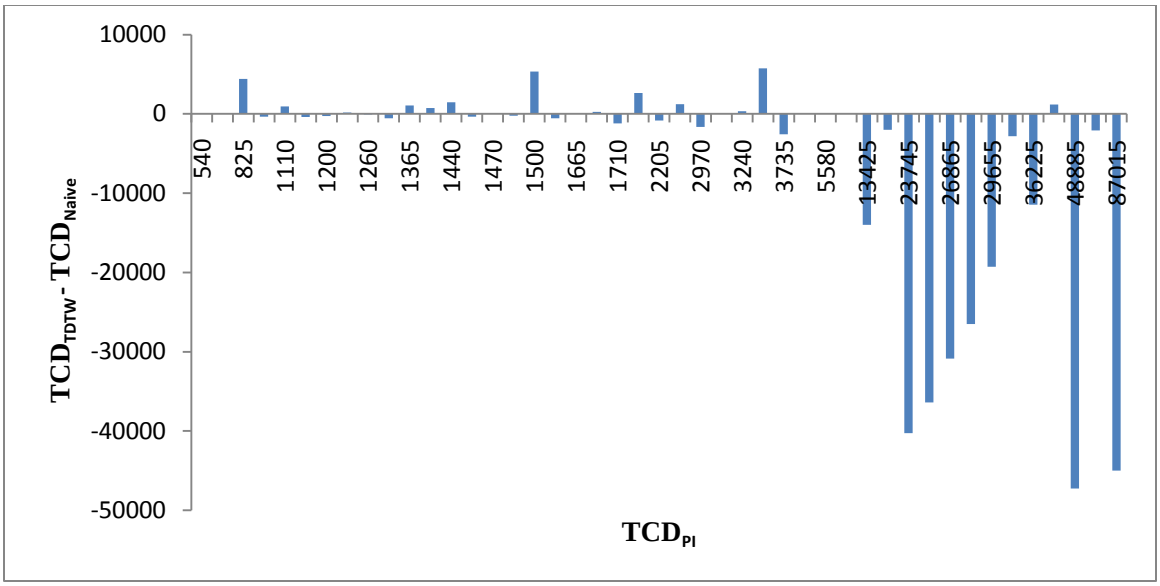


Figure 5.17 $TCD_{TDTW} - TCD_{Naive}$ versus $TCD_{PI}$ for ORD

## 5.6 Conclusions, limitation and future research

This thesis contributes towards the incorporation of day-of-operation weather forecast information to support probabilistic decision-making in the NAS. We provide several methodologies to generate day-of-operation probabilistic capacity scenarios from the day-of-operation weather forecast. These methodologies require the day-of-operation forecast, historical weather forecasts and historical realized capacity to generate the scenarios.

In Chapter 2, we qualitatively showed the influence of weather on arrival capacity. We discussed that strategic decisions, like ground delays, align the demand with future capacity. Stochastic GDP models determining optimal planned arrival rates by incorporating the uncertainty in future capacity have received attention but their application in a real world setting is limited. Previous research efforts have focused on development of these stochastic optimization models but the inputs required in the model have received little attention particularly the uncertainty in arrival capacity. Researchers assume that the uncertainty in arrival capacity can be obtained from weather forecast or from the expertise of air traffic managers. In this research we developed methods to model the uncertainty in arrival capacity in the context of making optimal ground delay decisions. Traditionally, the stochastic GDP models capture this uncertainty in arrival capacity through probabilistic capacity scenarios or as a Markov process. We opted to represent the uncertainty in the arrival capacity as probabilistic capacity scenarios over a Markov chain as the former are more tractable and have a direct application in current air traffic management. We showed that the current literature lacks methodologies on generating probabilistic capacity scenario generation from forecast forecasts. We stated that our contribution is to develop methodologies that incorporate the day-of-operation weather forecast for scenario generation.

In Chapter 3, we provided a discussion on the weather forecasts used in the research: Terminal Aerodrome Forecast and the STRATUS forecast. While STRATUS is forecast specific to SFO the TAF is a forecast issued for all major airports. Therefore the methodologies using the TAF can be applied to airport where this forecast is available. We assumed that the TAF issued between 4am and 6am and the STRATUS issued at 8am is used to make strategic decision for the entire day-of-operation. We then proceeded to develop TAF Clustering, a methodology involving clustering of TAF and followed by a clustering of realized arrival capacity profiles that have similar TAFs. Fog burn-off scenarios, a scenario generation methodology focusing on the SFO STRATUS forecast determines scenarios from day having similar fog burn-off time. This methodology is developed exclusively for SFO. We also showed that the current research lacks the methodology to assess the performance of the probabilistic capacity scenarios. To this end we developed a platform that allows us to measure the performance of the scenarios on the days in the past. Under this platform we used the SSGDM, a stochastic ground delay model, requiring probabilistic scenarios as inputs to determine the optimal planned arrival rates. The performance was based on average of the sum of ground delay cost and the realized air delay cost over 45 days in the past. Each day, the ground delay was obtained directly from the SSGDM. The realized air delay was determined using a queuing diagram between optimal planned arrival rate and the realized capacity profile of the day in the past.

Finally, to assess the benefit of weather forecast in decision making we formulated two scenario generation methodologies which serve as reference cases: Naïve Clustering and Perfect Information. Naïve clustering generates scenarios though clustering of arrival profiles without forecast information. Perfect Information assumes that the GDP can be formulated based on the precise information about future capacity. Comparing costs of scenario generation techniques requiring the weather forecast with Naïve Clustering quantifies the benefit of using weather forecasts in decision making. On the other hand, comparison with the cost under the Perfect Information measures the inaccuracy in weather forecasts.

We illustrated in Chapter 4, some of the short comings of TAF Clustering and motivated the development of another methodology involving the TAF called Dynamic Time Warping (DTW) Scenarios. We borrowed a technique from machine learning called DTW developed to match electrical signals. We used DTW to compare the day-of-operation TAF with all the TAFs in the past. While developing this scenario generation methodology, we introduced three parameters that tune DTW to improve comparisons between the TAFs. We wanted to determine the set of optimal parameter values that minimize the average delay costs. This was computationally expensive and we used an algorithm involving stochastic response surface methodology to determine the optimal values thereby maximizing the performance of DTW Scenarios.

In Chapter 5 we used the above mentioned methodologies to develop scenarios for four US airports: SFO, LAX, BOS and ORD. We found that incorporating TAF information to plan arrivals reduces the realized delay costs on average by 5%-30% when compared to methodologies that do not make use of a weather forecast. The average cost of delay using the scenarios developed from DTW Scenarios is the lowest across the four airports when compared to other scenario generation techniques. Moreover, the difference is cost is statistically different. This indicates that that DTW Scenarios can be applied to any airport where the TAF is issued. For SFO, our results indicate that the scenarios generated from STRATUS and TAF yield similar costs. One possible reason is the TAF for SFO is generated using the STRATUS forecast. Lastly,

the benefit of using weather forecast in decision is more pronounced on days that have a greater capacity-demand imbalance. This trend is clearly visible for SFO, BOS and ORD but the trend was not pronounced for LAX.

There are some limitations of the current research. First, the current research determines optimal ground delays by ignoring the en-route airspace. It assumes that the en-route airspace has infinite capacity and the only bottlenecks in the NAS are at the airports. In the present system, the air traffic controllers manage en-route traffic by issuing tactical decision including speed and heading changes to resolve conflicts after strategic decisions have been initiated. Furthermore, human factors like the increase in work load of controllers may contribute to additional delays. These additional sources of delays are not incorporated in the SSGDM. This makes it difficult to compare the observed ground delays in the NAS with the idealized delays determined from our platform. In future research we plan to develop a NAS wide ATFM model that accounts for uncertainty in capacity at several NAS resources. We can then use our platform to determine the delays associated with our strategic initiatives and compare them to the actually observed delays in the system.

Second, we use the TAF issued before 7am to make decisions for the entire day. The accuracy of the TAF might be lower for periods later in the day-of-operation. Similarly, STRATUS is a forecast product that is used to predict the burn-off time of the fog and the fog burns-off before 3:00pm almost surely. Therefore, planning of operations for periods beyond 3:00pm is outside the scope of the forecast. In future research we plan to reduce the planning horizon and then compare the realized costs using the TAF and STRATUS.

Third, the optimal parameter values and the average costs are determined using the same set of days in DTW Scenarios that may lead to over fitting of data. Ideally, the design should be determined from one set of days and the costs should be reported on another set of days. In this research we did not investigate whether the optimal parameter values relate to specific airport characteristics, particularly the arrival capacity and demand. In future research we plan to understand the relationship of the parameter values with arrival demand levels and capacity values.

Fourth, certain qualitative weather forecasts were ignored in the TAF while generating the scenarios. In future research we plan to include the qualitative forecast that might offer more insight into the scenarios and may lower the realized cost. Specifically for SFO, we plan to develop a methodology to generate probabilistic capacity scenarios by integrating the TAF and the STRATUS forecast.

## References


[1] C. Barnhart, M. Dresner, M. Hansen, K. Neels, A. Odoni, E. Peterson, L. Sherry, A. Trani, B. Zou M. Ball, "Total Delay Impact Study," NEXTOR, 2010.

[2] B. Zou, "PhD Thesis," University of California Berkeley, Berkeley, 2011.

[3] M. Wolfson, R. Hallowell, M. Moore B. Forman, "Aviation user needs for convective weather forecasts," in *American Meteorological Society 79th Annual Conference*, Dallas, TX.

[4] M. Wolfson, D Clark, S. Troxel, A. Madiwale, J. Andrews M. Weber, "Weather Information Requirements for Terminal Air Traffic Control Automation," in *American Meteorological Society 4th Int. Conf. on Aviation Weather,* Paris, France.

[5] NASDAC. (2000) NASDAC Review of NTSB Weather-Related Accidents. [Online]. http://www.asias.faa.gov/aviation_studies/weather_study/studyindex.html

[6] Journal of Ceiling and Visibility. [Online]. http://www.rap.ucar.edu/asr2002/j-c_v/j-ceiling-visibiltiy.htm

[7] Aviation Weather Center. (2002, February) Compendium of Ceiling and Visibility Articles for Operational Meteorology. [Online]. http://aviationweather.gov/static/info/cigvis.html

[8] G. Kulesa, "Weather and Aviation: How Does Weather Affect the Safety adn Operations of Airports and Aviation, and How Does FAA work to Manage Weather-related Effects," in *The Potential Impacts of Climate Change on Transportation*., 2003.

[9] A. Odoni R. de Neufville, *Airport Systems : Planning, Design and Management*.: The McGraw-Hill companies, 2003.

[10] Federal Aviation Administration. Air Service Performance Metric- Complete. [Online]. http://aspm.faa.gov/

[11] J. Foggin & G. D. Gosling W. Cotton, "Potential Future Contribution of Air Traffic Management Technology to the Capacity of San Francisco International Airport," in *EIR/EIS—Emerging and New Technologies*, San Francisco, 2001.

[12] M. Ball, C. Barnhart, G. Nemhauser, and A. Odoni, *Handbooks in operations research and management sciences: Transportation*.

[13] P.B. Vranas, D.J. Bertsimas, and A.R. Odoni, "The multi-airport ground-holding problem in

air traffic control," *Operations Research*, pp. 249--261, 1994.

[14] M. Helme, "Reducing air traffic delay in a space-time network," in *IEEE International Conference on Systems, Man and Cybernetics*, 1992, pp. 236-242.

[15] K. Lindsay, E. Boyd, and R. Burlingame, "Traffic flow management modeling with the time assignment mode," *Air Traffic Control Quarterly*, 1993.

[16] D. Bertsimas and S. Patterson, "The traffic flow management rerouting problem in air traffic control: A dynamic network flow approach," *Transportation Science*, vol. 34, no. 3, pp. 239-255, 2000.

[17] D. Bertsimas and S.Patterson, "The air traffic flow management problem with enroute capacities," *Operations Research*, vol. 46, no. 3, pp. 406-422, 1998.

[18] A.R. Odoni, "The Flow Management Problem in Air," in *Flow Control of Congested Network*. Berlin: Springer-Verlag, 1987, pp. 269 –288.

[19] Romanin-Jacur and G. Andreatta, "Aircraft flow management under congestion," *Transportation Science*, vol. 21, no. 4, p. 249, 1987.

[20] M. Terrab and A. Odoni, "Strategic Flow Management for Air Traffic Control," *OPERATIONS RESEARCH*, vol. 41, no. 1, pp. 138-152, January-February 1993.

[21] O. Richetta and A. Odoni, "Solving Optimally the Static Ground Holding Policy Porblem in Air Traffic Control," *Transportation Science*, vol. 27, no. 3, pp. 228-238, August 1993.

[22] O. Richetta and A. Odoni, "Dynamic solution to the ground-holding problem in air traffic control," *Transportation Research Part A: Policy and Practice*, vol. 28, no. 3, pp. 167-185, 1994.

[23] M.O. Ball, R. Hoffman, A.R. Odoni, and R. Rifkin, "A stochastic integer program with dual network structure and its application to the ground-holding problem," *Operations Research*, vol. 51, no. 1, pp. 167-171, 2003.

[24] A. Mukherjee and M. Hansen, "A Dynamic Stochastic Model for the Single Airport Ground Holding Problem," *Transportation Science*, vol. 41, no. 4, pp. 444-456, 2007.

[25] P. Liu, "Managing uncertainty in the single airport ground holding using scenario based and scenario free approaches," University of California- Berkeley, Berkeley, PhD Thesis 2007.

[26] A. Mukherjee and M. Hansen, "A dynamic rerouting model for air traffic flow management," *Transportation Research Part B: Methodological*, vol. 43, no. 1, pp. 159-171, 2009.

[27] A. Alonso, L.F. Escudero, and M. Teresa Ortuño, "A stochastic 0–1 program based

approach for the air traffic flow management problem," *European Journal of Operational Research*, vol. 120, no. 1, pp. 47-62, 2000.

[28] Yu-Heng Chang, "Stocastic Programming Approaches to Air Traffic Flow Management Under the uncertainty of Weather," Georgia Institute of Technology, Atlanta, PhD Thesis 2010.

[29] K. Hoyland and S. Wallace, "Generating scenario trees for multistage decision problems," *Management Science*, pp. 295-307, 2001.

[30] M. Kaut and S. Wallace, "Evaluation of scenario-generation methods for stochastic programming," *Stochastic Programming E-Print Series*, vol. 14, no. 1, 2003.

[31] G. Pflug, "Scenario tree generation for multiperiod financial optimization by optimal discretization," *Mathematical programming*, vol. 89, no. 2, pp. 251--271, 2001.

[32] N. Gulpinar, B. Rustem, and R. Settergren, "Simulation and optimization approaches to scenario tree generation," *Journal of Economic dynamics and control*, vol. 28, pp. 1291-1315, 2004.

[33] IBM Corporation, "IBM optimization library stochastic extensions users guide," 1998.

[34] P.B. Liu, M. Hansen, and A. Mukherjee, "Scenario-based air traffic flow management: From theory to practice," *Transportation Research Part B: Methodological*, vol. 42, no. 7-8, pp. 685--702, 2008.

[35] L.S. Cook and B. Wood, "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.

[36] Federal Aviation Administration, *Pilot's Handbook of Aeronautical Knowledge*.: Skyhorse Publishing, 2009.

[37] National Climatic Data Center. (2011, June) HDSS Access System,Station Selection. [Online]. http://hurricane.ncdc.noaa.gov/pls/plhas/HAS.FileAppSelect?datasetname=9957ANX

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

[39] L. Parsons, E. Haque, and H Liu, "Subspace Clustering for High Dimensional Data: A Review," *ACM SIGKDD Explorations Newsletter*, vol. 6, pp. 90-105, 2004.

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

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

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

[43] M. Vlachos, M. Hadjieleftheriou, D. Gunopulos, and E. Keogh, "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, p. 225.

[44] E.J. Keogh and M.J. Pazzani, "Derivative dynamic time warping," in *First SIAM international conference on data mining*, 2001.

[45] R.G. Regis and C.A. Shoemaker, "A stochastic radial basis function method for the global optimization of expensive functions," *INFORMS Journal of Computing*, vol. 19, pp. 457-509, 2007.

[46] D.R. Jones, "A taxonomy of global optimization methods based on response surfaces," *Journal of Global Optimization*, vol. 21, no. 4, pp. 345-383, 2001.

[47] San Francisco International Airport, "Weather and Operations at SFO: A primer for media," San Francisco, Media primer January, 2010.

[48] Mark Hansen and Tatjana Bolic, "Normalization of Airport and Terminal Area Operational Performance: A Case Study of Los Angeles International Airport," in *4th US and Europe,Air Traffic Management Research Seminar*, Santa Fe, New Mexico, 2001.

[49] MITRE Corporation, "Airport capacity benchmark report 2001," 2001.

[50] Federal Aviation Authority. BOS AAR. [Online]. http://www.fly.faa.gov/Information/east/zbw/bos/bos_aar1.htm

[51] Ricondo & Associates, Inc. , "O'Hare International Airport Master Plan," Chicago, 2004.