Estimating One-Parameter Airport Arrival Capacity Distributions for Air Traffic Flow Management

Tasha R. Inniss #, Michael O. Ball +
Department of Mathematics
Trinity College,Washington, D.C. 20117
InnissT@trinitydc.edu
+Robert H. Smith School of Business,
Institute for Systems Research
University of Maryland
College Park, MD 20742
MBall@rhsmith.umd.edu

August, 2002

Abstract

During instances of capacity-demand imbalances, efficient planning and decision-making in air traffic flow management is contingent upon the "goodness" of the capacity distributions that estimate airport capacity over time. Airport capacities are subject to substantial uncertainty as they depend on stochastic weather conditions. In this paper, we develop models that take into consideration the stochastic nature of weather. The main objective of this paper is the development of probabilistic capacity forecasts. To assess the improvements that could be gained by using the capacity probabilistic forecasts, the capacity distributions developed in this paper are input into existing static, stochastic, ground holding models, which uses probabilistic capacity forecasts and determines the amount of ground delay to assign to incoming flights.

1 Introduction

When an airport's arrival and/or departure capacity is reduced during "peak demand periods", demand for an airport's resources exceeds the capacity at which the airport can accept this demand. This is known as a capacity-demand imbalance. Demand refers to the number of flights scheduled to arrive or depart in a given time period (rate of flight arrivals or departures), whereas capacity is the maximum number of flight arrivals or departures per unit time. To address and manage this imbalance, traffic flow managers at the *Federal Aviation Administration*'s (*FAA*) *Air Traffic Control Systems Command Center* (*ATCSCC*), also known as the Command Center, may institute a ground holding procedure known as a *ground delay program* (*GDP*). During a GDP, flights are assigned delay to be taken on the ground at their departure airports until a time when they can safely arrive at their destination airports with little to no airborne delay.

Weather conditions at an airport determine which runway configurations and landing procedures are used. The combination of the runway configurations and the landing procedures determines an *airport's acceptance rate* (*AAR*) or operational capacity. There are 2 major types of landing procedures: *Instrument Flight Rules* (*IFR*) and *Visual Flight Rules* (*VFR*). Under VFR conditions at *San Francisco's International Airport* (*SFO*), aircraft normally arrive from the northwest in dual side-by-side approaches on runways 28L and 28R. See Figure 1 for the runway layout at SFO. When IFR conditions exist, the AAR is reduced because the landing of aircraft in pairs on the two closely spaced parallel runways is considered unsafe. Table 1 lists the capacities or AARs for the various combinations of runway configurations and landing procedures at SFO. VAPS is an acronym for visual approaches and has the same conditions as VFR with the addition of a ceiling that exceeds 3500 feet and a visibility that exceeds 7 miles at the San Mateo Bridge for SFO.

ftbpFU5.6991in2.7821in0ptRunway Layout at San Francisco's International Airport (Courtesy of the ATCSCC)sfo.bmpSince there is a direct relationship between weather conditions and an airport's acceptance rate or capacity through its runway configurations and landing procedures, accurate forecasts of weather conditions are crucial for an "good" estimation of airport capacity or AARs.

Overall airport capacity is comprised of two interdependent capacities, the arrival capacity and the departure capacity. The determination of an

Land	**Depart**	**IFR**	**VFR**	**VAPS**
28L 28R	1L 1R	30	45	60 (daylight); 50 (non-daylight)
28L 28R	28L 28R	30	45	45
28L or 28R	1L 1R	30	N/A	30
28L 28R	1L or 1R	30	45	45
1L 1R	1L 1R	30	N/A	30
19L 19R	10L 19R	27-30	N/A	45
19L 19R	19L 19R	25-30	N/A	42
19L or 19R	10L 10R	27-30	N/A	30
19L 19R	10L or 10R	27-30	N/A	45
10L 10R	10L 10R	27-30	N/A	37
Any Single	Runway	27	N/A	27

Table 1: AAR Chart for SFO (Courtesy of ATCSCC)

airport's overall capacity is a difficult task because weather conditions, runway configurations, arrival/departure ratios and the fleet (aircraft type) mix must all be considered. Eugene Gilbo, from the Volpe Transportation Systems Center, proposes methods in [8] and [9] for optimizing overall airport capacity by considering the complex relationship between arrival and departure capacities through an arrival/departure capacity curve. Hall [11] expands Gilbo's work by developing collaborative methods for allocating arrival and departure capacities. Though it is recognized that additional efficiencies can potentially be gained by considering both arrival and departure capacities simultaneously, arrival capacities and estimating arrival capacity distributions for a GDP will be the focus of this paper.

2 Background

Over the past several years, a new, collaborative process was conceived and implemented (January 1998) to enhance the decision-making processes in air traffic flow management. This new process, which is known as *Collaborative Decision Making* (*CDM),* was motivated by a need to combine information sources for increased information sharing and distributed decision-making. Prior to CDM, there existed a central planning paradigm in which the FAA's ATCSCC was viewed as a central planning authority that made decisions with little input from the airlines and their operational control cen-

ters (AOCs). Since the inception of CDM, there is a collaborative paradigm in which the airlines and AOCs have more control, flexibility and input into the air traffic flow management decision-making processes. Under CDM, GDPs have become more effective due to increased information exchange, more efficient tools and common situational awareness.

During a GDP under CDM (known as GDP-Enhancements or GDP-E), the ATCSCC sends out an advisory to the AOCs stating its intention to institute a GDP. Airlines then respond with any flight cancellations or delays. If there is still a need for a GDP, the ATCSCC assigns arrival slots to airlines based on their scheduled times of arrival. This process is known as *Ration by Schedule* (*RBS*). Each airline then has the option to redistribute its flights among the slots it has been assigned using the *substitution process*. Each airline may also cancel or delay additional flights. These changes are then submitted to the ATCSCC. If there are any unused slots (due to delays or cancellations), an algorithm known as the *Compression Algorithm* is used for inter-airline slot swapping. (For more detailed information on CDM procedures, see [6], [14], and [22]). CDM has made a significant positive impact on decision-making processes during GDPs. As of October 1999, there is a savings of over 4 million delay minutes due to the CDM procedure Compression. For more information on other benefits derived from CDM, see [4] and the *Free Flight Phase 1* (*FFP1*) Report [15].

A crucial aspect of the whole GDP-E process is determining the number of arrival slots per unit time. Determining the number of arrival slots or the amount of ground delay to assign to flights is known as the *ground holding problem* (*GHP*). Previous work have been done on the deterministic GHP ([5], [18]) and the stochastic GHP ([2], [3], [5], [10], [11], [13], [19], and [20],[21]). For the deterministic GHP, an airport's capacity is generated by the ATCSCC at the beginning of the day and is assumed to remain constant. This version of the GHP does not take into consideration the stochastic nature of weather. In the stochastic GHP, changing weather conditions are considered through the probabilistic forecasting of airport capacity or AAR. In this paper, we consider the problem of deriving probabilistic forecasts for the stochastic GHP.

We will focus on deriving probabilistic forecasts of arrival capacity for the Hoffman-Rifkin static stochastic ground holding model. The Hoffman-Rifkin model ([5], [13], [21]) was developed to be used in conjuction with CDM procedures, though it has not been adopted and is not currently being used by the ATCSCC. The model is formulated as an integer programming

problem and solved in polynomial time using linear programming (LP) relaxation techniques (since the model is a dual network flow problem). The inputs to the model are demand, which is assumed to be deterministic, the ground holding and airborne holding costs per flight, and the arrival capacity scenarios, which is the distribution of the AAR over time. Recall that AAR is the maximum number of flights that can be landed in a given unit of time. The output of the model is the *planned airport acceptance rate* (*PAAR*) or the number of flights assigned arrival times in each time interval. We can think of the output as the number of arrival slots per unit time. This model deals with flights in the aggregate and CDM procedures determine the slot assignments of individual flights. See the CDM GDP-E Concept of Operations diagram (Figure 2) ftpFU4.7478in5.8972in0ptProposed GDP-E Concept of Operationsconops.wmffor a graphical view of planning procedures during a GDP-E. The main contributions of this paper are the development of probabilistic capacity forecasts, which contain capacity scenarios needed by the Hoffman-Rifkin static stochastic GHP, and the modification of the Hoffman-Rifkin model to realistically represent dynamic changes in GDPs. In the next sections, we describe capacity scenarios and then derive probabilistic distributions of capacity scenarios.

3 Airport Arrival Capacity Scenarios

3.1 Conceptual Representation of Arrival Capacity Scenarios

On any given day, there is a weather forecast that translates into a particular capacity. As the weather (forecast) changes, so does the capacity. The severity of the weather and the accuracy of the forecast determine the amount of fluctuation in the capacity level. It is a normal practice for specialists at the ATCSCC to receive different weather forecasts from various sources. Each of these forecasts could realistically result in a different capacity scenario or *arrival capacity distribution* (*ACD*). The strength of the forecast could possibly determine the probability of a particular ACD.

We now describe, on a conceptual level, a range of possible ACD models. For modeling convenience, time is broken into discrete intervals, e.g. typical lengths are one-hour or 15-minute intervals. An ACD can be represented as a bar graph, where the x-axis represents time of day, y-axis represents arrival

capacity levels or number of flights able to land, and each bar represents arrival capacity over a given time interval.

In general, an ACD can take almost any structure imaginable. Thus the number of possible ACDs is enormous. In this section, we explore possible ACD structures. In the most general ACD model, there can be a constant fluctuation in the arrival capacity level, as seen in Figure 3. We shall refer to this as the "general" ACD because it can be used to model almost any given airport that may be plagued with constant weather or runway configuration changes.ftbpFU286.25pt177.875pt0ptForm of the General Arrival Capacity DistributionFigure A simpler model allows for only 2 capacity levels and the distribution fluctuates between these 2 levels. This type of model may adequately represent conditions at an airport with few runway configurations or with one main weather pattern that has multiple peaks throughout the day. This will be known as the "2-Level" ACD (Figure 4).ftbpFU253.125pt125.125pt0ptForm of the 2-Level Arrival Capacity DistributionFigure A further simplification of the 2-Level ACD model has a structure in which there is "normal" or maximum capacity at the beginning of the day, then reduced capacity for a certain length of time, followed by a return to the "normal" level. This model may capture most airports in which there is a consistent weather pattern that occurs sometime after sunrise and only lasts for a finite length of time. This ACD is appropriate for most airports whose arrival capacity level is cut almost in half when inclement weather forces a change from VFR to IFR approaches eliminating the possibility of closely spaced parallel approaches. In this case, we need only estimate the two parameters: start time and duration of reduced capacity. Thus, this is referred to as the "2-Parameter" ACD (Figure 5). ftbpFU244.125pt120.625pt0ptForm of 2-Parameter Arrival Capacity DistributionFigure The simplest ACD is a distribution that initially has reduced capacity, remains constant at this level for a given time and then increases to the normal arrival capacity level. Therefore, the only parameter to be estimated is the end time (duration) of the reduced capacity. Hence, this ACD model is the "1-Parameter" ACD (Figure 6). This is a reasonable way to model ACDs associated with weather patterns that are present at sunrise, remain continuously in place for a period of time and then clear.ftbpFU254.625pt133.4375pt0ptForm of 1-Parameter Arrival Capacity DistributionFigure In general, the task of estimating a vector of ACDs is quite daunting because of the possible model complexity of an individual (general) ACD. There is a range of model complexities, which can be seen

in the ACD models presented here. It was discovered that the structure of the two simplest models, the 2-Parameter and 1-Parameter ACDs, are representative of actual capacity scenarios for a reasonably broad range of airports. In fact, the 1-Parameter ACD can be applied in one very important and practical case. It can be used to model morning fog at San Francisco's International Airport (SFO). This case of modeling morning fog at SFO is the focus of this paper.

3.2 Estimating 1-Parameter Capacity Scenarios for SFO

In the case of early morning GDPs at SFO, morning fog exists at sunrise forcing IFR approaches and a planned AAR of approximately 30 flights per hour, as depicted in the top graph of Figure 7. When the fog burns off, the arrival capacity level returns to the "normal" value of 45 flights per hour (when visual approaches can be performed, See Table 1).ftpFU5.8807in3.2923in0ptIFR Capacity at SFO (Reproduced with Permission from MIT's Lincoln Laboratory)ifr2.wmfNominal IFR conditions at SFO are characterized by a ceiling of less than 2500 feet or a visibility of less than 3 miles. Since fog conditions are present at sunrise, the start of reduced capacity is at sunrise. To estimate the duration of reduced capacity, the duration of a GDP and the duration of IFR conditions given a GDP is planned will be calculated. Morning GDPs at SFO are planned to end at the burnoff time of fog or at the initial dissipation of the stratus conditions.[1] According to [10], stratus clouds form during overnight hours and dissipate during the morning hours. There are times when the dissipation or burnoff of the "cloudiness" occurs after the late morning arrival traffic peak (1800Z)[2] in which demand is high. During these times, there exists a capacity-demand imbalance due to fog. ftbpFU396.9375pt189.125pt0ptPercentage of GDPs Implemented During Morning Hours at SFOMorning GDPs at SFO were analyzed to ascertain the benefits of calibrating the models based on this set of data. The analysis is performed on GDP data from GDP logs from the ATCSCC for the years 1995, 1996, and 1997 at SFO. The GDP logs from the ATCSCC contain information on GDP parameters such as start time of the GDP, end time of the GDP, cancellation time of the GDP, AAR for the GDP, center(s) included

[1] From an e-mail correspondence with Forrest Terral at the ATCSCC on June 15, 1999.

[2] 1800Z refers to 1800 hours in Zulu time; at SFO, 10am PST. Zulu Time is also called Greenwich Mean Time (GMT) or universal time (UTC) and is based on the time at the zero degree meridian in Greenwich, England.

in the GDP and the maximum delay incurred during the GDP. In order to determine the duration of a GDP, we calculated cancellation time minus the start time for all morning GDPs during 1995, 1996, 1997. In Figure 8, we observe that more than 50% of the GDPs planned and implemented at SFO for any given month in any given year occur during the morning hours (1600Z-1900Z). Thus, restricting attention to morning GDPs at SFO is reasonable. Since fog materializes during times of low demand, weather data can be used to give an estimate of the duration of IFR conditions regardless of demand.

GDP durations on days when GDPs were planned and durations of nominal IFR conditions can be used for estimates of the durations of reduced capacity in 1-Parameter ACDs. Thus, there are two sources of data that can be used to estimate 1-Parameter ACDs for SFO. Since the Hoffman-Rifkin static, stochastic ground holding model requires a vector of ACDs, the vector will be determined by estimating a distribution of 1-Parameter ACDs.

3.3 Determining Distribution of 1-Parameter Capacity Scenarios

Our goal is to determine a distribution that is an estimate of the duration of IFR conditions during instances of capacity-demand imbalances for which a GDP will be implemented. In general, a distribution can be determined by binning a given set of observations or empirical data to create a relative frequency histogram. A frequency histogram is constructed by partitioning the range of observed values (largest value minus smallest value) into k equal length subintervals (bases of rectangles or bins) and by calculating the frequency (counts) of observations in each subinterval (heights of rectangles or bins). To construct a relative frequency histogram, simply divide the frequency in bin i, f_i, by the total number of data points, n. A relative frequency histogram estimates an underlying *probability distribution function* (*pdf*) because $\sum_i f_i/n = 1$. The 3 years of empirical GDP data from SFO are used to create the frequency histogram of duration of morning GDPs (Figure 9). This histogram is determined by considering the duration of GDPs conditioned on a GDP being planned.ftbpFU453.875pt222.9375pt0ptOverall Frequency Histogram for SFO Morning GDP Data Figure ftpFU445.5pt222.125pt0ptSmoothed Histogram for SFO Morning GDP DataFigure

Notice the peak between 4 and 5 hours in Figure 9. In the opinion of specialists, this is due to the operational procedures at the ATCSCC during

the given 3 years of data. There was a limit on the maximum number of hours a GDP could be run and that limit was 4 hours. To remove the effect of operational procedures on the overall distribution, a smoothing technique that smooth out peaks and valleys in a histogram is utilized. In this triangular binning technique, half of the number of data points in a given bin is added to one-quarter the number of data points in the bin on the (immediate) left side of the given bin and one-quarter the number in the bin on its (immediate) right side to give the new frequency or number of data points in the given bin. This process is repeated for all bins except the first and last bins. To account for end effects, smoothing of "end" bins is done by taking the weighted sum of three-quarters of the frequency in the given bin plus one-quarter of the frequency in the bin immediately next to the given bin. The smoothing of all bins result in a "smoothed frequency histogram" (Figure 10).ftpFU441.6875pt211.625pt0ptRelative Frequency Histogram for Duration of IFR ConditionsFigure

Each bin of a histogram corresponds to the duration of reduced capacity in a particular 1-Parameter ACD. Thus the histogram can be thought of as the distribution of 1-Parameter ACDs. To get an associated probability of an ACD (or bin on histogram), simply divide the frequency in a given bin by the total sum of frequencies (relative frequency):

$\mathcal{P}(S_i) = Frequency_i Sum of Frequencies$, for each scenario (ACD) S and bin i.

It can be argued that a distribution based on weather data is more representative of durations of IFR conditions. Distributions of IFR conditions at SFO were derived using data from "Surface Airways Hourly" from the *National Climatic Data Center* (*NCDC*) during the years 1984-1992. This database contains an hourly listing of surface weather observations that are taken at stations located primarily at major airports and military bases. These stations are operated by the *National Weather Service* (*NWS*), the U.S. Air Force Weather Service, the U.S. Navy Weather Detachment and the FAA. Data at these stations are collected using the *Automated Surface Observing System* (*ASOS*) that was designed primarily for aviation operations. The weather variables used to determine distributions of IFR conditions are ceiling height and visibility. Figure 11 gives the distribution, based on weather data, with the relative frequencies along the y-axis. ftbpFU422.875pt216.875pt0pt"Conditional" Distribution of Durations of IFR ConditionsFigure This distribution is a conditional distribution because it is the distribution of duration of IFR conditions given that the

duration equals or exceeds one hour. Thus, there is no zero bin in this distribution, but there is a zero bin on the GDP distribution. Since the goal is to estimate the duration of IFR conditions during instances of capacity-demand imbalances for which a GDP will be implemented, some reasonable combination of the two distributions is sought. One alternative is to derive a conditional distribution of the durations of IFR conditions given that a GDP is planned. To determine this "conditional" distribution, we simply include the zero bin from the GDP distribution in the distribution of IFR conditions and normalize accordingly to derive the new associated probabilities. See Figure 12.

Histograms are used to give information about an underlying *probability distribution function* (*pdf*) of empirical data. In this section, histograms created using ALL available empirical (GDP and weather) data were presented. The underlying pdf will be referred to as a *Capacity Probabilistic Distribution Function* (*CPDF*) and is the vector of 1-Parameter ACDs that will be used as input into the Hoffman-Rifkin model. The underlying CPDF is a distribution that is based on weather conditions that are highly stochastic in nature. It is feasible to think that the CPDF would change according to the changes in weather. Thus, the fundamental mechanism that controls the CPDF is continuously changing over time. A CPDF can be created from any given set of observations, in this case, for the set of years of available historical data at SFO. There are different types of CPDFs that can be derived by partitioning the overall CPDF into subunits based on the underlying changing mechanism (weather). In the next sections, models are presented that are the result of partitioning the overall CPDF in different ways.

4 Estimating (Types of) Capacity Probabilistic Distribution Functions

Given empirical data about capacity (or IFR conditions), relative frequency histograms can be constructed and used to estimate CPDFs. In order to account for changes over time in the underlying weather mechanism, daily, monthly or seasonal CPDFs are all types of distributions that could be of interest. The type that will ultimately be utilized depends on the operational preferences of the specialists at the ATCSCC, as well as other factors.

The specialists may be interested in using daily distributions, which could

be obtained by creating relative frequency histograms for a given day over many years of data. If there is not a substantial number of years of data, sample size becomes a problem. One way to address the problem of sample size for daily distributions is to group the daily data by month. This grouping would yield monthly CPDFs. It is possible that several months may have the same or similar CPDFs, especially in the case of distributions of IFR or inclement weather conditions. Weather conditions such as thunderstorms and snow occur at certain times of year or during specific seasons. At SFO, the most prevalent weather conditions are radiation fog and advection fog. Radiation fog is also known as ground fog and occurs when the temperature drops to the dew point near the ground. Advection fog occurs when warm, moist air moves over a colder land mass. According to the Weather Sensing Group at MIT's Lincoln Laboratory, radiation fog occurs more than 100 days annually and advection fog is the next most frequently occurring weather condition at SFO. (See Figure 13)ftFU6.2353in4.5109in0ptFrequency of IFR events at Major US Airports (Reproduced with Permission from MIT's Lincoln Laboratory)ifr3.wmf Based on conversations with specialists at the ATCSCC, fog is heaviest from September to the middle of March and burnoff times are difficult to ascertain. Through the Marine Stratus Initiative at SFO [7], which is led by the Weather Sensing Group at Lincoln Laboratory, it has been determined that the stratus cloud season is during the months of May to September. As an example, a possible CPDF may be the same during the months September-March and the same during the months May-September, but different for the two groupings of months. In the next section, we will present methods for determining seasonal CPDFs. It can be assumed that seasonal GDPs correspond to seasons of certain weather conditions. Thus, monthly distributions will be grouped into seasons, based on some measure of similarity, to create seasonal CPDFs.

4.1 Seasonal Distributions

Decomposing an overall CPDF into groupings of months (seasons) based on some measure of similarity (dissimilarity) is reminiscent of partitional clustering in which data is partitioned into disjoint clusters. This partitioning is done by minimizing a measure of dissimilarity within each cluster and maximizing the dissimilarity between different clusters. (See Hartigan [12] for more information on clustering techniques.) The data used for this paper are time ordered, so a simple clustering technique is not adequate. A method

is needed to perform clustering that is imbedded in a time series. This type of clustering will be referred to as "seasonal clustering". The resulting clusters must be contiguous and homogeneity should exist within the clusters. Since the data constitute a time series, a time plot could be used to detect seasonal trends as an initial step in determining seasonal clusters.

4.1.1 Detecting Seasonal Trends

Time plots can be used to gauge an overall trend in time-ordered data. Figure 14 is a simple time series plot of average GDP lengths from month to month for 1995 through 1997. Observe that the months October to March lie above the horizontal line (average GDP length = 3 hours) and the months May to September lie below the line. This observation gives credence to the idea that the CPDFs would differ between these two groupings of months, as stated in the previous section.ftbFU436.0625pt241.0625pt0ptTime Series Plot of Average GDP DurationsFigure In the next sections, model formulation and solutions of a more rigorous analysis to ascertain seasonal clusters are presented.

4.1.2 Model Formulation

Given the twelve months in a year, the goal is to partition the year into groupings of (contiguous) months that contain the most similar weather conditions. The problem of determining the optimal partitions (seasons) can be formulated as a set covering/partitioning integer programming problem. The goal of the set covering *integer program* (*IP*) is to "cover" the whole year by a finite number of covers or seasons with the smallest total cost. The goal of the set partitioning IP is to cover the whole year by a finite DISJOINT set of seasons in a least cost fashion. Recall that the set covering IP has the following formulation (See [16]):

$$\begin{aligned} \text{Minimize} \quad & \sum_{k=1}^{n} \mathrm{C}_k \mathrm{x}_k \\ \text{subject to} \quad & \sum_{k=1}^{n} \mathrm{a}_{jk}\ \mathrm{x}_k \geq 1, \text{ for each month } j \\ & \mathrm{x}_k \in \{0,1\} \end{aligned}$$

where A=[a_{jk}] is a 0-1 incidence matrix with $a_{jk} = 1$ if $j \in M_k$ (month j is in candidate season M_k), 0 otherwise; $\{M_k\}_{k=1}^{n}$ corresponds to the set of candidate seasons; n is the number of candidate seasons; $C_k$is the cost of including $M_k$in the cover; and x_k is a binary variable with value 1 when $M_k$is included in the cover, and 0 otherwise. (Recall that in the formulation for the set partitioning IP, the constraint is an equality.)

In this case of assigning months to seasons, the columns of A (i.e. the set of M_k's) can be efficiently enumerated since a season is characterized by a start month and an end month and the months must be contiguous. The possible seasons can be enumerated according to length of (contiguous) months. If all possible combinations are allowed, i.e. groupings 1 month in length (M_1,..., M_{12}), 2 months in length (M_{13},..., M_{24}) up to 12 months in length, there are a total of 133 possible seasons. Since there are 12 months, there are 12 different seasons for each possible season length except for the season of length 12 (only 1 way). Thus, 12 multiplied to 11 plus 1 results in 133 possible seasons. Intuitively, no weather season lasts more than 5 months. If the length of the season is restricted to being no more than 5 months, then there is a total of 60 possible seasons. Since there are 12 months and 5 different possible season lengths, enumerating the seasons yields 60 possible seasons in this candidate season set. Results will be given for the candidate season set of size 60. The incidence matrix A (for a candidate season set of size 60) is:

$$A = \begin{array}{c|cccccccccccc} & M_1 & M_2 & \ldots & M_{12} & M_{13} & M_{14} & \ldots & M_{24} & \ldots & M_{49} & M_{50} & \ldots & M_{60} \\ Jan & 1 & 0 & & 0 & 1 & 0 & & 1 & & 1 & 0 & & 1 \\ Feb & 0 & 1 & & 0 & 1 & 1 & & 0 & & 1 & 1 & & 0 \\ Mar & 0 & 0 & & 0 & 0 & 1 & & 0 & & 1 & 1 & & 0 \\ Apr & 0 & 0 & & 0 & 0 & 0 & & 0 & & 1 & 1 & & 0 \\ May & 0 & 0 & & 0 & 0 & 0 & & 0 & & 1 & 1 & & 0 \\ Jun & 0 & 0 & & 0 & 0 & 0 & & 0 & & 0 & 1 & & 0 \\ Jul & 0 & 0 & & 0 & 0 & 0 & & 0 & & 0 & 0 & & 0 \\ Aug & 0 & 0 & & 0 & 0 & 0 & & 0 & & 0 & 0 & & 0 \\ Sep & 0 & 0 & & 0 & 0 & 0 & & 0 & & 0 & 0 & & 1 \\ Oct & 0 & 0 & & 0 & 0 & 0 & & 0 & & 0 & 0 & & 1 \\ Nov & 0 & 0 & & 0 & 0 & 0 & & 0 & & 0 & 0 & & 1 \\ Dec & 0 & 0 & & 1 & 0 & 0 & & 1 & & 0 & 0 & & 1 \end{array}$$

Observe that the 0-1 incidence matrix, A, almost has the consecutive ones property; for example, the ones in columns M_{13}, M_{14}, M_{49} and M_{50} are

consecutive. A matrix has the consecutive ones property if in any given column, ones appear consecutively. Recall that a matrix having this property is totally unimodular (TU), and thus, the IP can be solved as an *linear program* (LP). This is a desired property because LPs can be efficiently solved using commercial software whereas IPs are, in general, more difficult. The consecutive ones property does not hold because there are wrap around columns such as M_{24} and M_{60} in matrix A. Though a matrix with consecutive ones and wrap around is not TU, it can be solved in polynomial time using a simple iterative procedure. This procedure involves rotating rows to delete the wrap around column and solving the LP for each rotation. The solution chosen is the best solution of all the optimal solutions of the rotations.

4.1.3 Seasonal Clusters Based on Average GDP Durations

In this section, a seasonal "clustering" technique that assigns consecutive months to a particular season based on some measure of similarity will be developed. A way to derive a finite number of seasons that contain contiguous months in a least costly fashion is desired. A set covering/partitioning integer program model will be used to determine the seasons.

Since the seasons are chosen in a least costly fashion, a cost of a season must be defined and determined. Conceptually, the cost of season M_j, C_j, is the "difference" between a month's CPDF and a season's CPDF. In this analysis, the cost function will be based on a difference in means. While this clearly represents an approximation, it should be noted that it appears that there exists a direct (increasing) relationship between the mean and the variance of GDP Durations (Figure 15). Hence, a cost function based on comparing means should also capture differences in variances, in this case.ftbpFU427.8125pt228.3125pt0ptRelationship Between Mean and Variance of GDP DurationFigure

Several cost functions are possible for comparing seasonal and individual monthly means. In this section, different cost functions will be given and compared. Section 4.3.5 will discuss a way to evaluate the quality of a given set of seasons based on a broader set of criteria.

The following cost functions are considered: (i) sum of squared deviations (SoSqs), (ii) normalized sum of squared deviations, or (iii) seasonal variances. The cost functions were chosen because they measure the difference between a season's mean and the means of the months contained in the season. The first cost is the sum of squared deviations between a season's value (average GDP

Sum of Squared Deviations(SoSqs)	$\sum_{j=1}^{m} (\overline{X}_{.j} - \overline{X}_{..})^2$
Normalized SoSqs	$()1m \sum_{j=1}^{m} (\overline{X}_{.j} - \overline{X}_{..})^2$
Seasonal Variances	$()1m - 1 \sum_{j} \sum_{i} (X_{ij} - \overline{X}_{..})^2$

Table 2: Seasonal Clustering Criteria (Cost Functions)

duration) and the values of the months contained within that season. The cost, normalized sum of squared deviations, is the sum of squared deviations divided by the number of months contained in that season. This cost function is chosen due to the possibly of a longer season being penalized by having a larger value for SoSqs. A seasonal variance is deemed appropriate because actual daily ground delay durations are considered. A seasonal variance is determined by calculating the variance of all daily GDP durations from the overall seasonal average. Table 2 gives the formulas for the three different clustering criteria. Here $\overline{X}_{.j}$ is the average over all days i in month j, $\overline{X}_{..}$ is the (overall) seasonal average over all days i and all months j, and X_{ij} is the GDP length on day i in month j.

As an example, the cost of the January/February GDP Season using SoSqs is calculated:

$$\begin{aligned} C_{13} &= (GDPAvg_{13} \text{ - } GDPAvg_1)^2 + (GDPAvg_{13} \text{ - } GDPAvg_2)^2 \\ &= (4.62 \text{ - } 5.08)^2 + (4.62 \text{ - } 4.16)^2 \\ &= .2116 + .2116 \\ &= .4232 \end{aligned}$$

where 13 denotes the January/February Season, 1 denotes month January and 2 denotes month February.

The candidate set of seasons must be enumerated and input into the set covering/partitioning model. Each season has a value: the average duration of a GDP in that season. For example, the value for January is the average of the Jan95 average GDP duration, Jan96 average GDP duration and Jan97 average GDP duration. The value for the Jan/Feb season is the averages of all GDP average durations for Jan95, Jan96, Jan97, Feb95, Feb96 and Feb97. It is possible that the set covering/partitioning procedure, under certain seasonal clustering criteria, could choose all seasons of length 1. Hence,

For 60 Possible Seasons:	N = 3	N = 4	N = 5
SoSqs	Apr-Jun Jul-Oct Nov-Mar	Apr-Jun Jul/Aug Sep/Oct Nov-Mar	Apr-Jun Jul-Sep Oct/Nov Dec/Jan Feb/Mar
Normalized SoSqs	Apr-Aug Sep/Oct Nov-Mar	Feb-Jun Jul-Sep Oct/Nov Dec/Jan	Apr/May Jun Jul/Aug Sep/Oct Nov-Mar
Seasonal Variances	Apr-Aug Jul-Oct Nov-Mar	Apr-Aug Jul-Oct Jul Nov-Mar	Mar-Jun Jul Aug Sep Oct-Feb

Table 3: Set Covering Solutions of GDP Seasons (n=60)

the following constraint, which limits the number of seasons chosen, is added to the set covering formulation:

$$\sum_{j=1}^{n} \mathrm{x}_j \leq \mathrm{N}$$

Note: N is the maximum number of covers or seasons.

To solve this set covering problem, the CPLEX Linear Optimizer 6.0 on a SUN Sparc10 Station was used. Table 3 gives the set covering solutions in terms of seasons for n=60. Observe for N=4, there is "over-covering" that occurs using the seasonal variance cost function. If set partitioning is used, then the resulting seasons (Mar-Jul, Aug, Sep, Oct-Feb) are disjoint. This approach seems more appropriate for a seasonal "clustering" method since the results of the set partitioning model ensures disjoint clusters.

It is interesting to note that the seasons determined by the SoSqs' seasonal clustering criterion correspond to the boxed seasons in the time plot of the monthly average GDP durations averaged over all 3 years, 1995, 1996 and 1997 (Figure 16). ftbFU408.1875pt200.4375pt0ptAverage of Monthly Average GDP Durations over 1995, 1996, 1997 Figure In the remainder of this paper, the seasons for this particular result will be referred to as: Winter GDP

GDP Durations in hours –¿	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**≥7**
Winter (Nov-Mar)	.21	.09	.07	.06	.15	.08	.08	.26
Winter-Smoothed	.14	.12	.08	.09	.12	.11	.07	.27
Spring (Apr-Jun)	.35	.15	.10	.12	.13	.06	.01	.08
Spring-Smoothed	.23	.21	.13	.13	.12	.07	.04	.07
Summer (Jul/Aug)	.35	.29	.17	.10	.05	.02	0	.02
Summer-Smoothed	.27	.30	.20	.12	.06	.03	.01	.01
Fall (Sep/Oct)	.21	.10	.27	.08	.12	.10	.08	.04
Fall-Smoothed	.14	.18	.19	.15	.11	.11	.08	.04

Table 4: ACD Probabilities for Frequency and Smoothed Histograms

Season (Nov/Dec/Jan/Feb/Mar), Spring GDP Season (Apr/May/Jun), Summer GDP Season (Jul/Aug), and Fall GDP Season (Sep/Oct). In Table 4, the relative frequencies or associated probabilities of ACDs in a particular season are given for both the frequency histograms and the smoothed histograms. Note that the probabilities that the duration of a GDP lasting 7 hours or more are higher during the "winter" season, as is expected since it takes longer for fog conditions to burn off during the winter months.

In this section, our analysis produced a partitioning of the months of the year into disjoint seasons, based on GDP data, with a corresponding distribution for each element in the partition. The same procedures could be applied using the weather data. In the next section, the clustering criterion developed will be based on differences in distributions rather than differences in means. This criterion will be applied to both the GDP data and the weather data.

4.1.4 Seasonal Clusters Based on Empirical Distribution Functions

In the previous section, CPDFs were based only on means due to the relationship between the means and variances. Since it is possible to have two distributions that have the same mean, but are different, the cost function (in this section) will be based on differences in distributions instead of differences in means.

An *empirical distribution function* (*EDF*), which is completely determined by observed values of a random variable, is used to estimate an underlying *cumulative distribution function* (*cdf*) of a group of observations or

empirical data. The *Kolmogorov-Smirnov* (*KS*) test is used to test if two or more samples come from the same distribution. Since the KS statistic measures the maximum deviation between the EDFs within classes and the pooled EDF, it will be used as the cost of a season in the cost function of the set covering/partitioning formulation. (See [17] for more detailed information on EDFs and KS tests.)

For any given season in the candidate season sets of size 60, an EDF is calculated for each month j in the given season according to:

$$F_j(x) = \frac{1}{n} \sum_{k=1}^{n} \mathrm{I}\{x_k \leq x\} \ , \ k = 1, ..., n$$

where n is the number of data points (days of GDPs or IFR conditions) in month j. For each real number x, $F_j(x)$ calculates the proportion of data that is less than or equal to that point x. The average of the monthly EDFs, known as the pooled EDF, gives the EDF for the season. The pooled EDF, $F(x)$ is computed by:

$$F(x) = \frac{1}{n} \sum_j n_j F_j$$

where n_j is the sample size for month j and $n = \sum_j n_j$. The KS statistic is appropriate for measuring the difference in a season's EDF and the EDFs of the months contained in that season. The KS statistic will be used as the cost of a given season in any of the candidate season sets and is calculated as:

$$\max_x \sqrt{\sum_j () n_j n [F_j(x) - F(x)]^2} \ , \ x = 1, 2, ..., n$$

A season whose KS statistic is small implies that the maximum deviation of any month's EDF from the seasonal EDF is small. Hence, the objective of the set covering/partitioning formulation is to minimize the maximum deviation of the months' EDFs from the seasonal EDF or minimize the KS statistic for a given season. Note that the KS statistic requires two or more classes (months) in order to be calculated, thus no single month seasons are allowed using this cost criterion. Table 5 contains the resulting seasons using both the GDP data and the weather data for 60 candidate seasons.

The seasonal clustering criterion (cost function) developed in this section have yielded many different sets of seasons. A method is needed to assess the quality of the sets of seasons and to determine which set is the "best" set.

For 60 Possible Seasons:	N = 3	N = 4	N = 5
Covering (GDP Data)	Apr/May Jun-Oct Nov-Mar	Apr-Jun Jul/Aug Sep-Jan Feb/Mar	Apr/May Jun-Oct Oct/Nov Dec/Jan Feb/Mar
Partitioning (GDP Data)	Apr/May Jun-Oct Nov-Mar	Apr-Jun Jul/Aug Sep-Jan Feb/Mar	Apr-Jun Jul/Aug Sep-Nov Dec/Jan Feb/Mar
Covering (Weather Data)	Mar-Jun Jul-Sep Oct-Feb	Mar/Apr May/Jun Jul-Sep Oct-Feb	Mar/Apr May/Jun Jul-Sep Oct/Nov Nov-Feb
Partitioning (Weather Data)	Mar-Jun Jul-Sep Oct-Feb	Mar/Apr May/Jun Jul-Sep Oct-Feb	Mar/Apr May/Jun Jul-Sep Oct/Nov Dec-Feb

Table 5: Set Covering/Partitioning Solutions using KS statistics (n=60)

4.1.5 Post Analysis for Evaluating Sets of Seasons

In previous sections, different cost functions yielded different sets of seasons. The question now is: which set of seasons is the "best" set? This section will discuss different methods for evaluating the quality of a given set of seasons. One way to evaluate a given set of seasons is by comparing the means of the different seasons to ascertain if they are statistically different from each other. This can be done using the method of multiple comparisons in a single-factor *analysis of variance* (*ANOVA*). Single-factor ANOVA is used to test whether there do indeed exist statistically differences in the means of the months [16]. The Single-Factor ANOVA model can be written as

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

where Y_{ij} represents the i^{th} observation of the j^{th} factor level

$$i = 1, ...n_j,\ j = 1, ..., k,$$

n_j is the number of observations for the j^{th} factor level, k is the total number of factor levels, μ is the overall mean of all factor level means, and α_j is called the effect of the j^{th} factor level. The unknown parameters (μ, α_j) are usually estimated from the data using the method of *ordinary least squares* (*OLS*). In OLS, $\sum_{j=1}^{k} \sum_{i=1}^{n_j} (Y_{ij} - \mu - \alpha_j)^2$ is minimized with respect to μ, α_1, α_2, ..., α_k.

Single-factor ANOVA must be performed before multiple comparisons because if there does not exist a difference in means (null hypothesis not rejected), then there is no need to determine where the differences are. An F-test is used to determine if there are statistically significant differences among the means of the months. Using the model involving the average GDP durations, the F-test tested the hypothesis that all factor level (monthly) means are equal and resulted in a p-value of .0030, which implies that there does exist some linear function of parameters that is significantly different from 0. In other words, there does exist a significant difference in means of the months.

It should be noted that the results of an F-test are valid only if certain assumptions about the error terms are satisfied. The error terms must be independent, have zero mean, constant variance (known as homoscedasticity), and must follow a normal distribution. In the case of daily GDP durations, the assumptions of constant error variance and normality were violated. The

	Contiguous Seasons	**Mean Square Ratio**
n=60	Mar-Jun vs Jul-Sep	14.06
	Jul-Sep vs Oct-Feb	24.39
n=133	May/Jun vs Jul-Sep	11.41
	Jul-Sep vs Oct-Apr	23.33

Table 6: Mean Square Ratios of Weather Seasons

assumptions on error terms are not often satisfied in practice, thus, caution should be taken with this method.

To avoid the issue of whether assumptions are satisfied or not in order for the F-test to be valid, the mean square ratio (better known as the F-value if normality is satisfied) can be used to evaluate a given set of seasons. The mean square ratio is the ratio of the mean square between groups (seasons) and the mean square within groups (seasons). It is desired to have seasons that exhibit homogeneity within seasons and variability between seasons. A mean square ratio that is large confirms that this is the case. The mean square ratios are computed for pairwise contiguous seasons. If the minimum of these values is greater than some large constant, e.g. 10, then the set of seasons is valid. The set of seasons resulting from the set partitioning procedure that minimized the KS statistic using weather data satisfy the mean square ratio criterion. See Table 6. For the candidate season set of size 60, the season Mar-Jun will be referred to as the "Rainy/Transition" Season, Jul-Sep as the "Summer Weather" Season and, Oct-Feb as the "Heavy Fog" Season. The CPDFs for these seasons will be used as input into the Hoffman-Rifkin model and results will be compared with Command Center plans.

Many cost functions were given in previous sections to determine sets of seasons and the post analysis in this section is used to determine the best set of seasons. Best refers to a set of seasons where there exists as much intra-season homogeneity as possible and as much inter-season variability as possible.

5 Decision Models for Determining Assigned Ground Delay

5.1 The Hoffman-Rifkin Static Stochastic Ground Holding Model

In their theses, Ryan Rifkin (Massachusetts Institute of Technology)[21] and Robert Hoffman (University of Maryland)[13] developed integer programming models to address the static stochastic version of the *ground holding problem* (*GHP*). (For a succinct description of the model, see [5]). Recall that the GHP is the problem of determining an optimal balance between the amount of delay to assign to flights to be taken on the ground during a GDP and the amount of expected airborne delay. The *Hoffman-Rifkin* static stochastic ground holding model (*H-R*) [5] is formulated as:

$$\text{Minimize} \quad \sum_{t=1}^{T} c_g G_t + \sum_{q=1}^{Q} \sum_{t=1}^{T} c_a p_q W_{q,t}$$

subject to

$$A_t - G_{t-1} + G_t = D_t \qquad t = 1, ..., T+1 \qquad (1)$$

$$G_0 = G_{T+1} = 0$$

$$-W_{q,t-1} + W_{q,t} - A_t \geq -M_{q,t} \qquad t = 1, ..., T+1, \qquad (2)$$

$$q = 1, ..., Q$$

$$W_{q,0} = W_{q,T+1} = 0$$

$$A_t \in Z_+, W_{q,t} \in Z_+, G_t \in Z_+ \qquad (3)$$

The objective of the H-R model is to minimize the sum of the costs of assigned ground delay and the costs of expected (unplanned) airborne delay. The decision variables, A_t, represent the number of flights that should arrive at the airport in time period t with no airborne delays. One can think of A_t as a planned AAR (PAAR) during time period t. Also, $W_{q,t}$ is the number of flights delayed in the air from time period t until a subsequent time period under scenario q and $M_{q,t}$ is the arrival capacity during time period t under scenario q. A sequence of $M_{q,t}$ for the whole time horizon, T, is one possible

capacity scenario q or ACD. Recall the various forms of ACDs as discussed in Section 3.1. The interpretation of G_t has certain subtleties. It can be thought of as the number of flights delayed on the ground from time period t to $t+1$. However, here time is measured relative to the time at which these flights would arrive at the airport. The actual time at which the delay would be taken is determined by choosing the specific flights to be delayed and then subtracting the appropriate (flight-specific) en-route times. Constraint set (1) states that all flights that are predicted to arrive in time period t (demand or D_t) or were delayed on the ground from the previous time period (G_{t-1}) should arrive in the current time period (A_t) or be delayed on the ground until a subsequent time period (G_t). Constraint set (2) states that under scenario q, all flights scheduled to arrive in the current time period or that are air delayed from a previous time period ($W_{q,t-1}$) must be air delayed until a subsequent time period or must arrive in the current time period.

The inputs into the H-R model are: the number of predicted arrivals or demand for each time period t (D_t), the cost of ground delaying one flight for one time period (c_g), the cost of one period of airborne delay of a single flight (c_a), and Q capacity scenarios (ACDs) with associated probabilities, p_q. The output of the model is the number of flights that should land in a given time period t, A_t, i.e., the number of arrival slots that should be made available in each time period t.

The H-R model assumes that all ground delay assigned under a particular output scenario is realized, deterministic and independent of the scenario. The model does not take into consideration the dynamic changes that may occur if one scenario is planned and another occurs. Thus, the model does not give the flexibility to make changes in assigned ground delay as forecasted weather conditions change.

5.1.1 Results using Seasonal CPDFs

The capacity scenarios (1-Parameter ACDs) and their associated probabilities derived in Section 4.1.5 are required inputs into the H-R model, along with the demand or predicted arrivals for each time period over a given (discretized) time horizon, the cost of one unit of (assigned) ground delay, and the cost of one unit of expected airborne delay. In this section the ACDs for SFO will be used as inputs into the H-R model to determine assigned ground holds. The H-R model will then be modified to allow for dynamic revisions to the assigned ground holds and results from the modified H-R model will

be compared to actual ATCSCC plans.

To compare the results of the modified H-R model to ATCSCC plans, actual GDPs during 1998 were run through the H-R model to determine the PAAR from the model (M_PAAR) over a given time period. *Aggregate Demand Lists* (*ADL*) were used to determine the aggregate demand for each time period 4 hours in advance of the planned start time of the GDP. Specialists at the ATCSCC plan a GDP at least 4 hours in advance based on forecasted weather conditions, predicted demand and capacity. In the ADLs, there are only 7 periods (hours) of predicted demand 4 hours in advance of the proposed start time of the GDP. Due to this constraint in the ADLs, each ACD contains only 7 periods of planned capacity from the proposed start time of the GDP. There are a total of 8 input capacity scenarios. The associated probabilities used depend on the seasonal CPDFs of choice. For analysis purposes, the seasonal CPDFs used are those resulting from the set partitioning method that minimized differences in EDFs implemented on weather data. These seasons were referred to as the "Heavy Fog" (Oct-Feb) season, the "Rainy/Transition" (Mar-Jun) season and the "Summer Weather" (Jul-Sep) season. These seasonal CPDFs were chosen for analysis because the F-values between contiguous seasons satisfy the mean square ratio criterion for a good set of seasons (Section 4.1.5). The associated probabilities for the 1-Parameter ACDs in the seasonal CPDFs (Figures 17, 18,19) are conditioned appropriately for the inclusion of the 0-hour reduced capacity ACD from the GDP data. The costs of one unit (minute) of ground delay and air delay were based on a study by the Air Transport Association (March 2, 2000) and reported by Metron, Inc. [1] The study concluded that the cost of one minute of delay at the gate is \$24.30, the cost of one minute of taxi-out delay is \$30.47 and the cost of one minute of airborne delay is \$47.64. Based on these values, one unit of airborne delay costs 1.96 times more than one unit of ground delay. The H-R model was run with $c_g = 1$ for three alternative airborne delay factors: $c_a = 1.5, c_a = 2.0$, and $c_a = 2.5$. Note the most realistic representation of c_a is 2.0 since airborne delay is 1.96 times more costly than ground delay. The M_PAAR (PAAR resulting from H-R model) results are given in Table 7.

ftbpFU164.3125pt174.8125pt0ptCPDF for "Heavy Fog" SeasonFigure ftbpFU161.3125pt174.06
for "Rainy/Transition" SeasonFigure ftbpFU159.0625pt166.5625pt0ptCPDF
for "Summer Weather" SeasonFigure

Weather Season	$\mathbf{c}_a = 1.5$	$\mathbf{c}_a = 2.0$	$\mathbf{c}_a = 2.5$
Oct-Feb	3 hours	4 hours	5 hours
Mar-Jun	2 hours	3 hours	4 hours
Jul-Sep	2 hours	2 hours	3 hours

Table 7: Results of the H-R Model (Number of Hours of Reduced Capacity)

5.1.2 Observations and Limitations of the H-R Model

It was empirically observed that the resulting plan (PAAR), or capacity scenario, output by the H-R model always corresponded to one of the input ACDs, i.e. the optimal solution to the H-R model always coincides with that of a deterministic problem that uses one of the capacity scenarios in the probabilistic forecast as inputs. (The conditions under which this property is guaranteed to hold will be explored in a subsequent paper). However, determining which scenario should be used, nonetheless is a problem that must be solved. Assuming the output scenario will always correspond to one of the input scenarios, there are a finite and small number of possible scenarios that could be "cost out" to determine the best scenario. The H-R model assumes that ground delay is deterministic and does not change as the weather changes. It can be shown that if inclement weather dissipates, some assigned ground delay can be recovered. On the other hand, if inclement weather persists longer than anticipated, additional delay may be incurred. In this latter case, the H-R model assumes that all additional delay is in the form of airborne delay. This is not necessarily the case as will be seen in the next sections. Thus, the H-R model overestimates airborne delay. Because of all the aforementioned observations and limitations of the H-R model, a modified version of the H-R model is proposed in the next section.

5.2 General Decision Model

The H-R model attempts to capture the stochastic nature of weather through the probabilistic distribution of capacity scenarios. It outputs a particular scenario based on demand, the air to ground cost ratio and the probability of the scenario. It does not capture the existing ability to dynamically change assigned ground delay as (predicted) conditions change. For example, if a dissipation of predicted weather occurs, then it may be possible to reduce previously assigned ground delay. Thus, some assigned ground delay may

be recovered if a GDP is canceled due to dissipation of inclement weather. Alternatively, if the duration of poor weather is longer than expected, then the GDP can be revised/extended, hereby assigning additional ground delays. In the next sections, we will discuss how to modify the H-R model so that its cost functions realistically reflect the dynamic updates that take place during GDPs as mentioned above.

5.2.1 Adjusting Assigned Delay in Canceled GDP

Here we consider the case where a GDP is canceled, i.e. the duration of reduced capacity is less than the value specified in the PAAR. Depending on how far a flight's (controlled) departure time is from the cancellation time of a GDP, the flight can recover some or all of its assigned ground delay. Suppose a flight f had an *original estimated time of departure* ($OETD$) of 12:30, but under a GDP, it was given a *controlled time of departure* (CTD) of 1:15. Now suppose that the inclement weather clears such that the GDP is canceled at 12:15. Since flight f is still on the ground at this cancellation time and full capacity has been restored, it is allowed to take off as soon as possible. However, a number of factors, e.g., the status of the passengers, might delay the time at which the flight is able to depart. For example, its *actual runway time of departure* ($ARTD$) might be 12:45. In this case, some of its assigned delay (CTD-OETD) is recovered. The actual *ground delay* (GD) realized is ARTD-OETD. In this scenario:

- Assigned GD = CTD - OETD = 1:15 - 12:30 = 45 minutes
- Actual GD = ARTD - OETD = 12:45 - 12:30 = 15 minutes
- GD Recovered = Assigned GD - Actual GD = CTD - ARTD = 1:15 - 12:45 = 30 minutes

Flights whose CTDs are prior to the GDP cancellation time (CNXTime) will incur all of their assigned GD. Thus, only flights that are controlled to depart after the CNXTime can recover some of their assigned GD. If a flight's OETD is before the CNXTime, then the amount of assigned GD that is available for recovery is CTD-CNXTime. (See Figure 20).ftbpFU6.8398in3.8865in0ptGround Delay Available for Recovery (Recoverable GD)cnxplot2.wmfThus, the percentage of assigned recoverable GD that is actually realized is:

$$ARTD - \max(OETD, CNXTime)\min(CTD - OETD, CTD - CNXTime)$$

Time Intervals	**% GD Recovered**
0-30 mins	0 %
31-60 mins	40.80 %
61-90 mins	65.20 %
91-120 mins	77.15 %
$\geq$ 120 mins	100 %

Table 8: Average Percentages of GD Recovered in a Canceled GDP

The above percentage can be greater than 1 if a flight's ARTD is greater than its CTD. This means that the flight incurred extra delay (possibly) unrelated to the GDP. In our data analysis, such flights are assumed to have incurred 100% of their assigned GD. To determine the percentage of recoverable GD that was recovered, we subtract the above value from 1:

$$1 - [ARTD - \max(OETD, CNXTime)\min(CTD - OETD, CTD - CNXTime)]$$

Using the information in the ADL files for all flights scheduled to arrive at SFO on all days in 1998 that a GDP was planned and run during the morning hours, the percentage of recoverable GD recovered as a function of a flight's CTD minus CNXTime is calculated.

Several methods for filtering the data were implemented. No one filtering method resulted in a set of data to which a good function could be fitted (low R^2) due to noise. Thus, averages of percentages of GD recovered are calculated in 30-minute intervals. It is assumed that any flight whose CTD is no more than 30 minutes after the CNXTime recovers none (0%) of its assigned GD and any flight whose CTD is more than 2 hours after the CNXTime recovers all (100%) of its assigned GD. Table 8 gives the average percentages for each 30-minute bucket **(**except for the $\mathbf{0-30}$ mins and the $>$ **120** mins buckets**)**. These assumptions will be observed and the respective percentages will be used to modify the output of the H-R model. The amount of assigned (recoverable) GD in a canceled GDP is adjusted by the (recoverable) amount that is recovered,

Recoverable GD realized =
Assigned Recoverable GD - (% GD Recovered)*(Assigned Recoverable GD).

5.2.2 Adjusting Assigned Delay in Revised GDP

With the emergence of CDM came the flexibility to "revise" different parameters of the GDP as conditions (weather or demand) change. If a GDP is revised/extended due to the worsening of weather conditions, then the originally assigned ground delay is modified and, thus depends on the scenario. The H-R model assumes that ground delay is deterministic and that, if reduced capacity lasts longer than the duration in the planned scenario, then all "extra delay" is in the form of airborne delay. Thus, it overestimates airborne delay. See Figure 21. If a flight's CTD is before the revised time (RevTime) of the GDP, then it may indeed incur unplanned *airborne delay* (AD), but if the flight's CTD is after the RevTime, then it should incur only extra GD.

In a GDP, flight delays are initially calculated by setting a *controlled time of arrival* (CTA). Assigned GD is set equal to CTA-OETA, and GD is then added to the OETD to determine the CTD. Flights are exempted from the revised portion of the GDP if

CTD - RevTime ¡ 0

and the airborne delay they may incur is calculated by subtracting their CTAs under the planned scenario from their CTAs under the actual scenario. ftFU404.875pt144.9375pt0ptAdditional GD under Revised GDPFigure On the other hand, flights whose CTDs satisfy

CTD - RevTime ¿ 0

can be assigned additional delay on the ground. The additional delay is calculated just as stated above for the other case, but all the delay is taken on the ground. Hence, assigned GD from the H-R model needs to be adjusted appropriately for the flights in a revised GDP.

5.2.3 Procedure for Comparing Planned and Actual Capacity Scenarios

In this section, the results of the H-R model, modified as described in the last two subsections for the actual GDPs in 1998, will be compared to the actual Command Center's plans for those same GDPs. Let F be a set of flights (f) scheduled to arrive at a congested airport. We denote by M_PAAR the

PAAR based on results from the H-R model, by CC_PAAR the PAAR based on the Command Center's planned duration of the GDP, and by ACT_PAAR the PAAR based on the actual duration of the GDP (baseline).

Given the planned and actual durations of reduced capacity, the procedure assigns CTAs or virtual arrival slots to the flights that fall into these durations or thereafter. The assigning of virtual slots is a basic recursion procedure and thus, can be viewed as a deterministic queuing model. If the AAR is 30 flights per hour, the slots are uniformly distributed over the hour interval. The procedure proceeds by calculating the amount of assigned ground delay ($CTA^f - OETA^f$) under each "plan", adjusting the amounts appropriately depending on whether the plan is greater than the actual or vice versa. Total Assigned GD under M_PAAR (TGD_M) is determined by TGD_M = $\sum_{f \in F} \text{GD}^f_{M\_PAAR}$. Total Airborne (weighted) Delay (TAD_M) under M_PAAR for Revised/Extended GDP is determined by TAD_M = ($\sum_{f \in F} \text{AD}^f_{M\_PAAR}$)*$c_a$.(Similarly, total assigned GD and weighted AD can be calculated under CC_PAAR.) Total Weighted Delay Under M_PAAR (TWD_M) is calculated by summing total GD and total weighted airborne delay. (Similarly, total weighted delay can be computed for CC_PAAR.) Average Total Weighted Delay is calculated over a representative sample of GDP Days in 1998 (see next section for discussion) under M_PAAR:

$$\text{AvgTWD\_M} = \frac{1}{n} \sum_{d=1}^{n} Sum_d(TGD\_M, TAD\_M).$$

(Similarly, average total weighted delay can be computed for CC_PAAR.) Average total weighted delay under M_PAAR and CC_PAAR are compared to ACT_PAAR according to:

AvgTWD_M - Avg(TGD_ACT) compared to AvgTWD_CC - Avg(TGD_ACT).

5.2.4 Computational Results of Comparisons of H-R Results to Command Center Plans

We now describe the results of experiments that tested the models and algorithms described in the paper. The "representative" GDP days used in the analysis of M_PAAR and CC_PAAR are days in 1998 whose ADL files did not contain any unreliable data due to temporary lapses in the data stream over the CDMnet and whose initial PAAR is 30 flights per hour. Due to these restrictions, there were not many GDP days available for analysis. Therefore

AvgTGD_M = 7284	*AvgTGD_CC = 8914*	*Avg(TGD_ACT) = 6875*
AvgTAD_M = 2417	AvgTAD_CC = 1314	Avg(TAD_ACT) = 0
AvgTWD_M = 9007	AvgTWD_CC = 9850	Avg(TWD_ACT) = 6875

Table 9: Total Weighted Delay of H-R Plans Vs Command Center Plans

a sample that represented the overall outcomes of all GDPs in 1998, was chosen. This representative sample is based on the breakdown of the types (outcomes) of GDPs in 1998 at SFO. (See Figure 22) The representative sample includes 11 GDPs, of which 6 are canceled, 4 are revised/extended and 1 is run out. Run out means that the original planned duration of the GDP is the same as the actual duration of the GDP. To determine M_PAAR, the reduced capacity level used was 30 flights per hour (since this is IFR level for SFO), the airborne delay cost factor (c_a) used was 1.5, and the demand levels used were the actual demands on the GDP days. CC_PAAR are the actual durations that were PLANNED for the GDPs in 1998. Act_PAAR are the durations that actually occurred on the GDP days.

Average total weighted delay was determined for M_PAAR, CC_PAAR, and ACT_PAAR according to the procedure outlined previously. Table 9 gives a summary of these results.

ftbpFU319.375pt174.0625pt0ptPercentage of Outcomes of 1998 Morning GDPsFigure According to Table 9, both the average of total GROUND delay and average of total weighted delay under M_PAAR are less than these values under the Command Center's plan. The H-R model produces less total weighted delay than the Command Center's plan and this is confirmed because the difference between the average total weighted delay under M_PAAR and the average total (ground) delay under ACT_PAAR is less than the difference between the average total weighted delay under CC_PAAR and the average total (ground) delay under ACT_PAAR:

$$AvgTWD\_M - Avg(TWD\_ACT) = 9007 - 6875 = 2132 \text{ minutes, and}$$

$$AvgTWD\_CC - Avg(TWD\_ACT) = 9850 - 6875 = 2975 \text{ minutes.}$$

Overall for the representative sample of GDP days in 1998, there could have been a savings of 843 delay minutes if our model had been used. Since there is a reduction in delays, our model is capturing what it should. In general, our model results in shorter planned durations of GDP than is currently employed. It appears that the best recommendation to the ATCSCC is to

plan shorter programs since there is a reduction in delay minutes. Since the current practice at the ATCSCC is to assign a reduction of capacity for a fixed duration regardless of the season, our approach seems more effective since it takes into consideration the weather season and assigns duration of reduced capacity accordingly.

6 Conclusions

With the tremendous growth of air traffic demand and congestion of the airspace comes an increase in the need for innovative methodologies and improved decision-support tools for effective *Collaborative Traffic Flow Management* (*CTFM*). Our analyses have indicated possible policy changes (by suggesting the use of shorter GDPs). In addition, it generates the required input for stochastic GDP models, which have the potential to substantially improve existing decision-support tools. The research in this paper creates a foundation on which further research can be built for the development of models that would increase the effectiveness of CDM procedures. The obvious next step would be to apply the techniques in this paper to model weather conditions at relevant airports using either the 1-parameter of 2-parameter ACDs.

Currently, new decision-support tools, such as Collaborative Convective Forecast Product (CCFP) and Integrated Terminal Weather System (ITWS), have been developed to enhance the quality, accuracy and timeliness of weather forecasts. A dynamic version of our approach could be developed by feeding output from a terminal area weather model into the CPDF generation process.

References

[1] "Aircraft Operating Costs by Stage of Flight and Associated Air Traffic Control Delay Costs Based on a Typical Flight,"1999, *Air Transport Association*, Washington, D.C., 2000.

[2] Andreatta, G., and Romanin-Jacur, G. (1987) "Aircraft Flow Management Under Congestion," *Transportation Science*, **21**, 249-253.

[3] Andreatta, G., Odoni, A.R. and Richetta, O. (1993) "Models for the Ground-Holding Problem," in *Large-Scale Computation and Information Processing in Air Traffic Control*, L. Bianco and A.R. Odoni, eds., Springer-Verlag, Berlin, 125-168.

[4] Ball, M., Hoffman, R., Hall, W. and Muharremoglu, A. (1998) "Collaborative Decision Making in Air Traffic Management: A Preliminary Assessment," Technical Report RR-99-3, NEXTOR, UC Berkeley.

[5] Ball, M., Hoffman, R., Odoni, A., and Rifkin, R. (1999), "The Static Stochastic Ground Holding Problem with Aggregate Demands," Technical Report RR-99-1, NEXTOR, UC Berkeley.

[6] Ball, M., Chen, C-Y., Hoffman, R., Vossen, T. (1999) "Collaborative Decision Making in Air Traffic Management: Current and Future Research Directions," *Proceedings of ATM '99, Workshop on Advanced Technologies and Their Impact on Air Traffic Management in the 21st Century.*

[7] Clark, D.A. and Wilson, F.W. (1996) "The Marine Stratus Initiative at San Francisco International Airport", *Project Report ATC-252*, MIT Lincoln Laboratory, Lexington, MA.

[8] Gilbo, E.P. (1993) "Airport Capacity: Representation, Estimation, Optimization," *IEEE Transactions on Control Systems Technology,* **1**, 308-313.

[9] Gilbo, E.P. (1997) "Optimizing Airport Capacity Utilization in Air Traffic Flow Management Subject to Constraints at Arrival and Departure Fixes," *IEEE Transactions on Control Systems Technology,* **5**, 490-503.

[10] Glockner, G.D. (1996) "Effects of Air Traffic Congestion Delays Under Several Flow Management Policies," Ph.D. Dissertation, Georgia Institute of Technology.

[11] Hall, W. (1999) "Efficient Capacity Allocation in a Collaborative Air Transportation System," Ph.D. Dissertation, Massachusetts Institute of Technology.

[12] Hartigan, J.A. (1975) *Clustering Algorithms*, John Wiley and Sons, New York.

[13] Hoffman, R. (1997) "Integer Programming Models for Ground-Holding in Air Traffic Flow Management," Ph.D. Dissertation, University of Maryland.

[14] Hoffman, R., Ball, M., Hall, B., Odoni, A.R. and Wambsganss, M. (1999) "Collaborative Decision Making in Air Traffic Flow Management," Technical Report RR-99-2, NEXTOR, UC Berkeley.

[15] Knorr, D., Wetherly, J., Ball, M., Hoffman, R., Inniss, T., Tanino, M., Shisler, L., Brennan, M., Blucher, M., DeArmon, J., Darrow, A.M., Howard, K., Gilbo, E., Danz, R., Biros, J. (1999) "An Operational Assessment of Collaborative Decision Making in Air Traffic Management: Measuring User Impacts through Performance Metrics," Document Control Number: R90145-01, FFP1, Federal Aviation Administration.

[16] Nemhauser, G.L. and Wolsey, L.A. (1988) *Integer and Combinatorial Optimization*, John Wiley and Sons, Inc., New York.

[17] Neter, J. and Wasserman, W. (1974) *Applied Linear Statistical Models*, Richard D. Irwin, Inc., Homewood, Illinois.

[18] Odoni, A.R. (1987) "The Flow Management Problem in Air Traffic Control," In Flow Control of Congested Networks, A.R. Odoni, L.Bianco, and G. Szego, eds., Springer-Verlag, Berlin, 269-288.

[19] Richetta, O. and Odoni, A.R. (1993) "Solving Optimally the Static Ground-Holding Policy Problem in Air Traffic Control," Transportation Science, **27**, 228-238.

[20] Richetta, O. (1995) "Optimal Algorithms and a Remarkably Efficient Heuristic for the Ground-Holding Problem in Air Traffic Control," *Operations Research*, **43**, 758-770.

[21] Rifkin, R. (1998) "The Static Stochastic Ground-Holding Problem," Master's Thesis, Massachusetts Institute of Technology.

[22] Wambsganss, M. (1996) "Collaborative Decision Making Through Dynamic Information Transfer," *Air Traffic Control Quarterly*, **4**, 107-123.

[23] Wilson, F.W., and Clark, D.A. (1997) "Estimation of the Cost of Ceiling and Visibility Events at Major Airports," *7th Conference on Aviation , Range, and Aerospace Meteorology*, Long Beach, CA.