An integrated model of flight and passenger delay for policy analysis in the National Air Transportation System

by

He Sun

B.E., Tsinghua University (2012)
Submitted to the Department of Civil & Environmental Engineering
and Department of Electrical Engineering & Computer Science

in partial fulfillment of the requirements for the degrees
of
Master of Science in Transportation
and
Master of Science in Electrical Engineering and Computer Science
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 2016

Author..... Signature redacted .....
Department of Civil & Environmental Engineering
Department of Electrical Engineering & Computer Science
May 19, 2016

Certified by..... Signature redacted .....
Amedeo R. Odoni
Professor of Civil and Environmental Engineering
Thesis Supervisor

Certified by..... Signature redacted .....
Dimitri Bertsekas
McAfee Professor of Engineering
Thesis Supervisor

Accepted by..... Signature redacted .....
Heidi Nepf
Donald and Martha Harleman Professor of Civil Engineering
Chair, Graduate Program Committee

Accepted by... Signature redacted .....
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

An integrated model of flight and passenger delay for policy analysis in the National Air Transportation System

by

He Sun

Submitted to the Department of Civil & Environmental Engineering
and
Department of Electrical Engineering & Computer Science
on May 19, 2016, in partial fulfillment of the
requirements for the degrees of
Master of Science in Transportation
and
Master of Science in Electrical Engineering and Computer Science

Abstract

Demand for air travel has increased over the years and so have airport delays and congestion. Delays have a huge impact on airline costs and influence the satisfaction of passengers, thus becoming an important topic of research in the field of air transportation. In recent literature, a Passenger Delay Calculator (PDC) was proposed to estimate passenger delays. The PDC computes passenger delays for a specified day based on actual flight schedules, fight cancellation information, and ticket booking information. However, since actual fight schedules are a necessary input, the PDC cannot be applied directly to hypothetical scenarios, in which different cancellation strategies are implemented and their impact on passenger delays are evaluated. A different model. Airport Network Delays (AND), has also been developed recently. The AND model estimates fight delays and relies on an input in which demand consists of the national planned fight schedule for any given day. In this thesis, we have attempted to incorporate these two models, the AND and the PDC, within a single framework, so that the resulting new integrated model can compute passenger delays without requiring an actual flight-schedule input. The integrated model would certainly increase the usefulness and applicability of the PDC since it could be used with hypothetical scenarios, different flight cancellation strategies, etc.

We first describe the framework of the integrated model for studying flight delays and passenger delays at a daily scale. The integrated model includes four components: a Tail Recovery Model, Flight Cancellation Algorithms, a Refined Airport Network Delay (RAND) model, and the PDC. The Tail Recovery Model recovers missing tail numbers for many flights recorded in the Aviation System Performance Metrics (ASPM) database. The Flight Cancellation Algorithms implement alterna-

tive strategies for flight cancellations in the presence of large delays, such as cancelling flights with long flight delays or flights with a large ratio of flight delay divided by the seating capacity of the aircraft. The RAND model is an extension of the AND, in which two implicit assumptions of the AND model have been modified. The RAND model produces better estimates of flight delays in the sense of replicating actual flight delays obtained from the ASPM database.

The overall integrated model is able to compute passenger delays and relies only on planned flight schedules rather than actual flight schedules. Moreover, the integrated model facilitates the study of factors that influence flight delays, such as weather conditions and demand fluctuations, and evaluates the impact of different cancellation strategies on passenger delays. Using actual data from different days, we conclude that passenger delays can be reduced on the busiest traffic days through improved flight cancellation strategies.

In the second part of the thesis, we extend the RAND model to compute flight delays on a monthly scale using different capacity profiles as input. These capacity profiles can be directly obtained from Federal Aviation Administration (FAA) reports or constructed by using classical machine learning algorithms on airport-level data. We validate our estimation of flight delays by using data of January, 2008, showing that both the capacity profiles and the RAND perform well in terms of replicating the actual monthly flight delays. These results imply that an effort can be made to develop an integrated model incorporating the RAND, the PDC etc. at a monthly scale or even at any generic time scale.

Thesis Supervisor: Amedeo R. Odoni
Title: Professor of Civil and Environmental Engineering

Thesis Supervisor: Dimitri Bertsekas
Title: McAfee Professor of Engineering

Acknowledgments

I would like to express my tons of thanks to my thesis supervisor Prof. Amedeo Odoni for guiding and supporting me all these years. You have set an example of excellence as a mentor, friend, and grandfather. Your discussion, ideas, feedback, and even editing have been absolutely invaluable. As one of your last students before you retire, I feel lucky and honorable.
I would like to thank my thesis reader Prof. Dimitri Bertsekas for supporting my second master degree in Electrical Engineering and Computer Science at MIT. As the last graduate student and two times Teaching Assistant working with you before you retire is my unforgettable honor.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

According to the Bureau of Transportation Statistics (BTS), the total performed aircraft departures by U.S. air carriers has increased from 5.5 million in 1985 to 9.3 million in 2012 (a 70% increase). One of the problems caused by this enormous growth is the emergence of large delays and congestion throughout the National Airport System (NAS). Before the 1990s, delays occurred only in a few airports in the NAS. However, as airport operations increased in the 1990s, the number of scheduled flights approached the capacity limits of many airports. As a result, delays and congestion became a widespread problem affecting network-wide operations. According to the Aviation System Performance Metrics (ASPM) database, the average flight delay in the U.S. increased from 10.1 minutes in 2002 to 16.7 minutes in 2007 (a 65% increase)[1]. The number of flights with taxi-out time exceeding 1 hour [2] increased from 17,331 in 1989 to 39,388 in 2013 (a 127% increase).

Delays and congestion also cause flight cancellations and passenger disruptions, leading to a huge economic loss to the airlines. According to the U.S. Congress Joint Economic Committee, in 2007, this cost was estimated to be $19 billion. Therefore, it is increasingly important to develop reliable models for analyzing delays and to propose policies for mitigating their impacts.

In the literature, delays are evaluated according to two types of metrics: flight-based

[1]This statistic is based on the Individual Flight Module of the ASPM. See Section 2.1 for detailed information.

[2]This statistic is based on the Individual Flight Module of the ASPM.

and passenger-based. Flight-based metrics focus on evaluating on-time performance of flights in terms of the difference between their arrival (departure) scheduled time and their actual arrival (departure) time; passenger-based metrics stress the impact of delay on passengers in terms of the difference between their planned arrival time and the actual arrival time. Since computing flight delays requires only flight itinerary information, whereas computing passenger delays requires additional information such as booking data and flight cancellations, different models are developed for estimating flight delays and passenger delays.

In this chapter, we introduce flight delays and passenger delays. We will then review the Airport Network Delays model (AND)[Pyr12] and the Passenger Delay Calculator (PDC)[BFV10] which are used to evaluate these two types of delays, respectively. These two models are essential cornerstones of our new integrated model that will be introduced in Chapter 2. Section 1.4 presents the research goals, as well as the contribution of this research. The final section contains an outline of the structure and the content of Chapter 2 and Chapter 3.

1.1 Flight delays

A flight is delayed when the realized arrival (departure) time of a flight is later than the scheduled arrival (departure) time. The difference between the two is identified as the flight delay associated with that specific flight. Aggregating individual flight delays, we are able to study flight delays at a network scale, for example, for all the flights departing from or arriving at any U.S. airport.

The U.S. Department of Transportation (DOT) defines a flight as a delayed one if it has more than 15-minute delay. Those with smaller delays are defined as on-time flights. This metric is referred to as the 15OTP metric and it is commonly used in the literature in the study of flight delays. Shumsky [Shu95] and Hall [Hal99] reported that US carriers increase planned block times and scheduled gate-to-gate times in order to improve the on-time performance in terms of 15OTP. This practice

results in a greater operating cost. Caulkins [Cau93] examined the trade off between the 15OTP on-time performance and operating cost, and proposed an approach to estimating airline's schedule reliability, which compares the performance of a specific airline to the average performance of all the airlines.

In the most recent twenty years, flight delays at airports in the U.S. have become a significant problem. According to the Bureau of Transportation Statistics, every year since 1990, roughly 20% of the flights of the largest U.S. Air Carriers were delayed with respect to the 15OTP metric[3], with the figures peaking in 2000 and 2007, at 27.4% and 26.6%. Figure 1-1 shows the evolution of average arrival delays from 1989 to 2013(ASPM)[4]. For each year, the average arrival delay is more than 10 minutes and it increases from 10.1 minutes in 2002 to 16.7 minutes in 2007. Figure 1-2 shows the evolution of the number of enplaned revenue passengers and the number of performed aircraft departures by U.S. Carriers, where the orange line represents the Enplaned Revenue Passengers and the blue line represents the number of of flights performed. These two measures of demand present a similar trend to the average flight delay. Before the 1990s, demand was relatively small, and delays were not as widespread. The busiest year in history in terms of number of flights was 2007. After 2007, the economic crisis reduced the number of flights and, as a result, airport delays. The Federal Aviation Administration (FAA) also introduced slot controls at New York's airports, which reduced the number of airport operations per hour, thus leading to a gradual reduction of demand and delays.

Flight delays at the network level can be explicitly described through the flight delays at each individual airport. Figure 1-3 shows the average arrival delays for 34 major U.S. airports in 2007 and 2013[5]. For 2007, the busiest year in the history of the

[3]The largest U.S. Air Carriers are those that take up at least 1% of the total domestic scheduled service

[4]The average arrival delay for a specific year is computed by averaging over arrival delays of all the flights in that year. For flights arriving on time or in advance, we set arrival delays to zero

[5]The list of 34 major airports with their abbreviation is shown in Table A.1. This set of 34 airports is identical to the set of the 35 busiest commercial airports in the country, known as the Operational Evolution Partnership (OEP 35) airports with the exception of Honolulu International Airport(HNL). The OEP 35 set includes all of the large hubs and several medium hubs in the U.S.

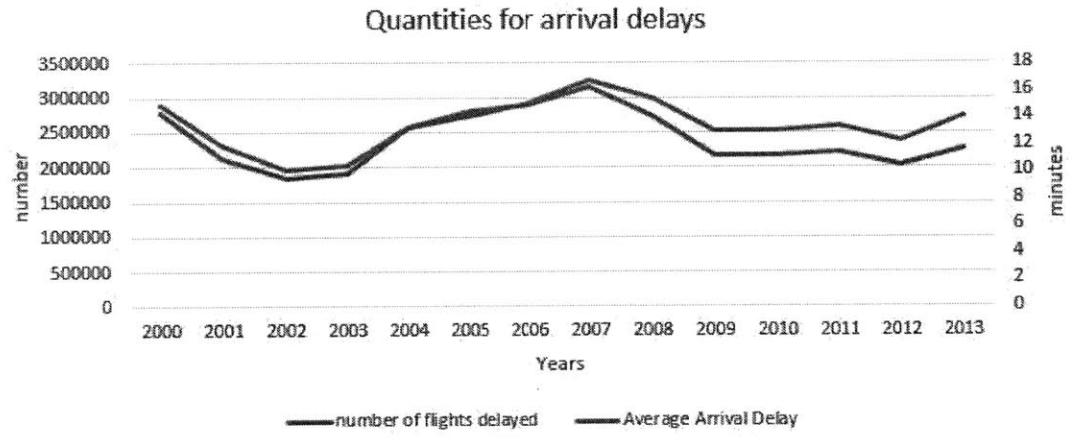


Figure 1-1: Number of scheduled flights delayed and average arrival delay for all scheduled flights in the US from 2000 to 2013

U.S. airline industry, the average arrival delay for most of the major airports exceeded 15 minutes, while the three major airports of the New York Metroplex–Newark Liberty Airport(EWR), LaGuardia Airport (LGA) and John F. Kennedy Airport (JFK) had average delays of more than 25-minutes.

At an individual airport, flight delays are caused by the relation between the actual demand and the capacity of the airport. As a day proceeds, flight delays fluctuate according to the evolution of demand and airport capacities. In particular, demand and capacities can be represented by the number of scheduled flights and the airport processing rates, respectively. In Figure 1-4, we relate these two factors to flight delays by presenting the evolution of the number of scheduled flights over time of day in five of the busiest U.S. airports in 2007–Atlanta (ATL), Chicago (ORD), Miami (MIA), Dallas/Fort Worth (DFW) and Newark (EWR). The average arrival delays are computed for each hour of a day as the average of the arrival delays associated with flights scheduled to arrive within that specific hour.

From Figure 1-4[6], we observe that from 1am to 6am, the average delay less than 15 minutes. After 6 am, the average arrival delays reach their peaks. At night, the

[6]To compute average accumulated arrival delays in a day, we first divide a day into 24 hours. For each hour, we compute the total arrival delays for flights scheduled to arrive within that hour. Then we sum up total arrival delays for all days in a year, and take the average over the number of flights, ending up with the quantities shown in this figure.

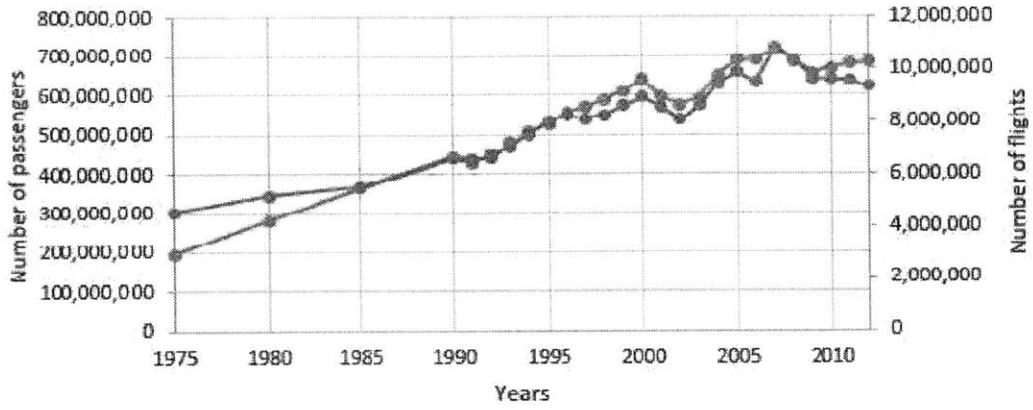


Figure 1-2: Scheduled flights and the number of passengers flown in the US from 1970 to 2012

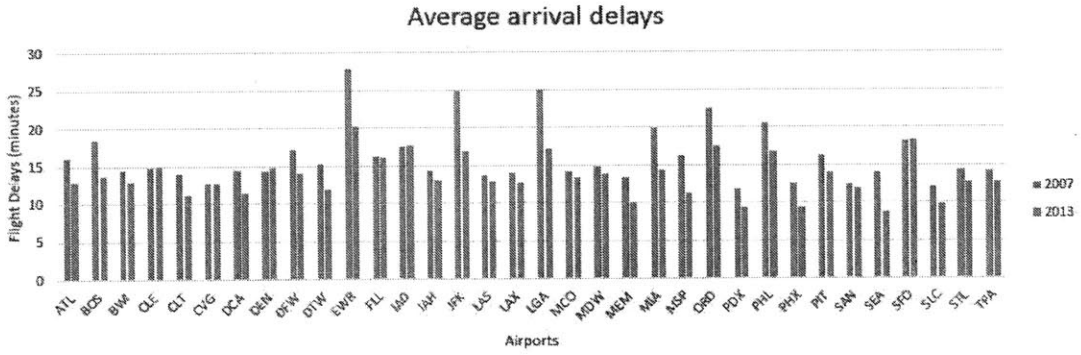


Figure 1-3: Average arrival delays of 34 major airports in the US for 2007 and 2013

delays fall until 1 am of the next day except for DFW.

The flight delays at an individual airport are also affected by the propagation effect from other airports in the network. In particular, when an upstream airport is congested due to bad weather conditions, all the flights through it are delayed. These delays may eventually propagate to all downstream airports. Therefore, when some major airports are very busy in a day, all the airports in the network will feel the effect and the flight delays throughout the network may be huge.

Due to this network propagation effect, when studying delays at an individual airport, we consider flight delays as a combination of propagated delays and local delays. Figure 1-5 shows the local arrival delays and total arrival delays for 34 major airports

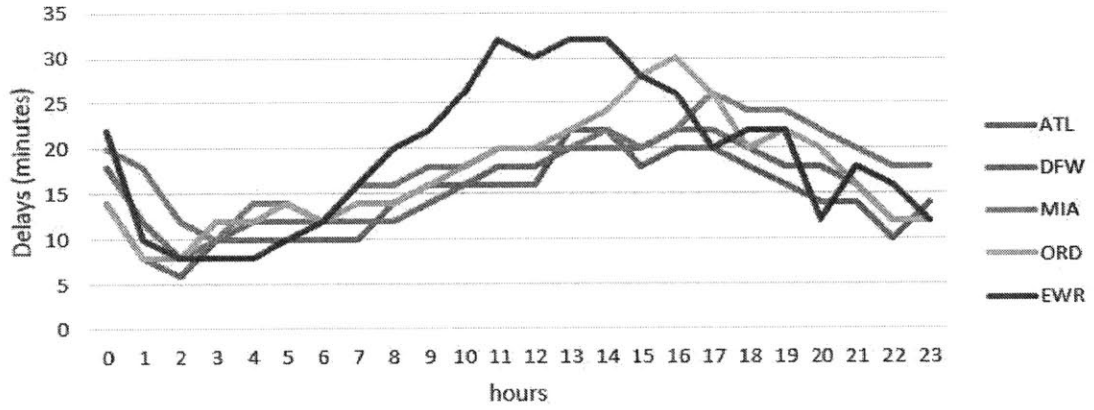


Figure 1-4: Evolution of actual average accumulated arrival delays over time of day

in the NAS modelled by the AND model. This computation takes place for June 14th, 2007, which was a day with light load. For such days, it can be seen that local delays may take up only a small portion of total delays, while the remaining delay is captured by propagation delays.

Various models have been introduced to capture the propagation of delays over the

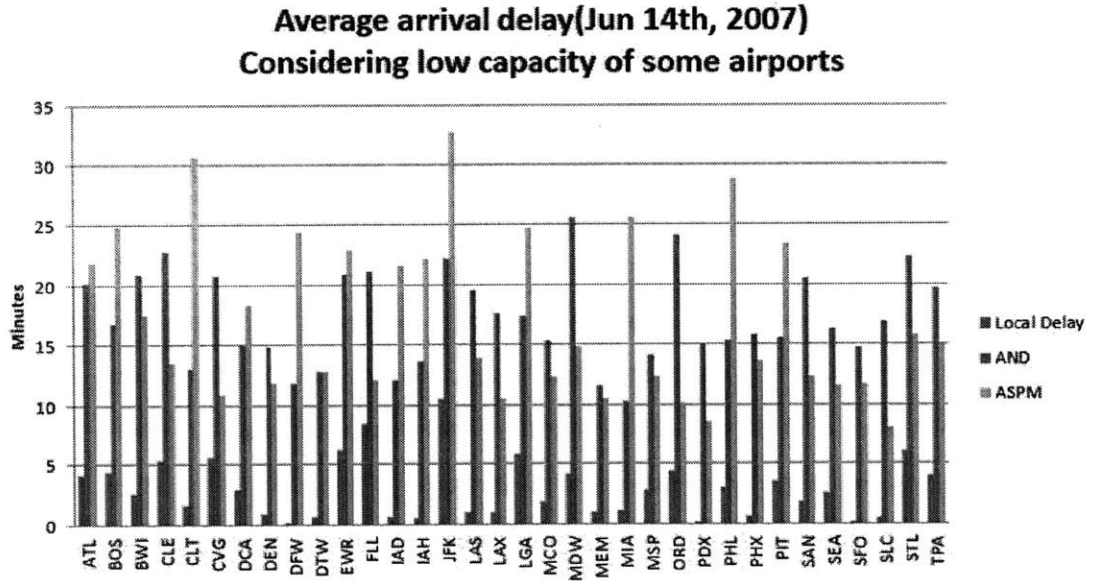


Figure 1-5: Average Local Arrival Delays and Average Total Delays from AND versus Delays from ASPM

course of a day and over the entire network of airports. An analytical model approx-

imating flight delays for an individual airport was introduced by Kivestu[Kiv74]. It models the take-off and landing process as a dynamic and stochastic queue. Following this work, Pyrgiotis[Pyr12] proposed the AND model, which computes the propagation of flight delays at a network scale.

1.2 Passenger delays

Several existing models provide reasonably accurate estimates of flight delays. However, practically all existing models do not consider the impact of flight delays on passengers. In fact, the difference between flight delays and passenger delays can be significant. For example, in 2007, the average passenger delay was estimated to be 33.09 minutes[BFV10], while the average flight delay was 16.69 minutes. This difference is due to several factors that contribute to passenger delays: the seating capacities of aircraft, flight cancellations and missed connections between flights.

Because missed connections and flight cancellations are not considered by models of flight delays, Bratu proposed the concept of passenger-based metrics for delays[Bra04]. Passenger delays measure the difference between the planned and actual arrival time at the final destinations of passengers' itineraries. To illustrate flight delays and passenger delays, we use the following example. Suppose an aircraft of American Airlines (AA) is scheduled to depart from Chicago O'Hare Airport (ORD) at 8:35am and arrive in Phoenix Airport(PHX) at 11:00am. However, due to bad weather conditions and airport congestion at ORD, the flight's actual departure time is changed to 11:28am, eventually changing the actual arrival time to 1:55pm. Thus, this flight has 173 minutes departure delay and 175 minutes arrival delay. Continuing this scenario, if another aircraft of AA is scheduled to depart from PHX at 11:40am for Los Angeles Airport (LAX), connecting passengers from Chicago will miss their connection at Phoenix because the aircraft from ORD to PHX will arrive after 11:40. These passengers will have to be transferred to a different flight to L.A.,but they will have

to wait for flights having available seats, thus increasing their travel time. Thus, passenger delays occur, which consist of both the flight delays and the waiting time for an alternative connecting flight. Besides missed connections and flight delays, flight cancellations also play a role in passenger delays and sometimes have an even greater impact. For example, assuming the initial flight is cancelled in Chicago, all the passengers on this flight must be reallocated to other flights and all the itineraries of these passengers will be disrupted. Therefore, it is very important to analyze passenger delays in addition to flight delays. By exploring the effects of different cancellation strategies under hypothetical scenarios, one can understand the impact of these strategies on passenger delays and provide suggestions for reducing passenger delays and increasing the total social welfare in the entire network.

1.3 Literature review

In this section, we review the AND model and the PDC. Our integrated model is based on these two.

1.3.1 Airport Delay Network Model

The Airport Network Delay model (AND) [Pyr12] consists of a numerical queueing model (QE) and a propagation delay algorithm (DPA), which capture the local delay effect of an individual airport and the propagation delay effect of the network, respectively. It estimates flight delays for each individual flight and tracks the itinerary of each aircraft during its life cycle of a day. By aggregating such flight delays at the individual airport level, statistics that are of interest, such as the average arrival delay for each airport, can be computed and used to analyze the performance of each airport. In addition, as a byproduct, the estimated arrival and departure times of each individual flight will be obtained. The AND requires scheduled flight information, airport capacities, and turnaround times of flights as its inputs. The model is very flexible in the sense of allowing the testing of many hypothetical scenarios. By changing the inputs, we can investigate the corresponding impact of different factors

on flight delay statistics, and have a sense of which factors have an important role or act as a bottleneck for the system. The AND describes the evolution of flight operations over the day by dividing a day into many small time intervals, running delay estimation for each individual flight within each interval, and updating information on each flight in terms of its estimated departure and arrival time. Schematically, Figure 1-6[Pyr12] shows each component of the AND model and the relations among these components. By default, the AND divides a day into time slots of 15 minutes

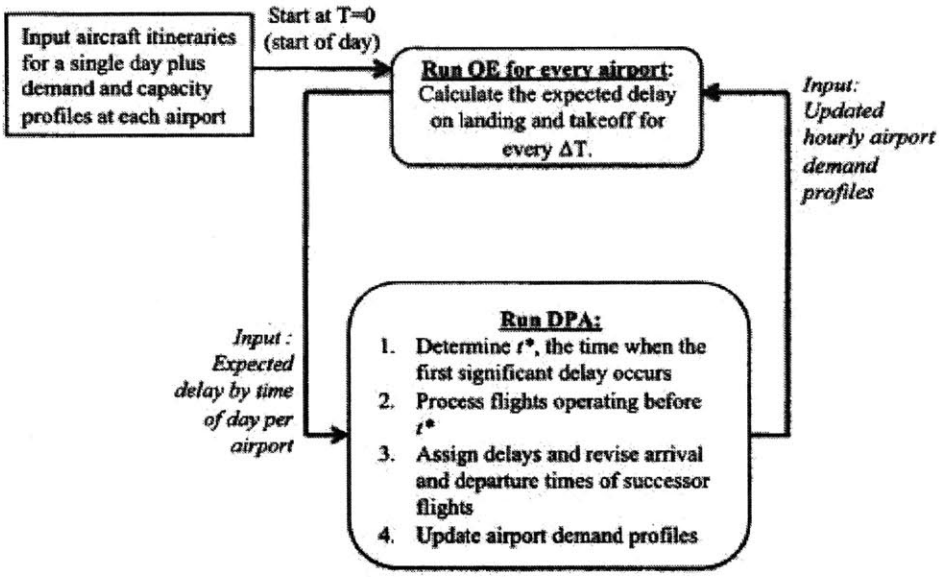


Figure 1-6: AND framework

and iterates between the QE and the DPA within each time slot. In the beginning of the loop, the scheduled flight information for all the flights of the day are input to the QE. The QE will compute the estimated delays for flights that land or take off within each time interval and update their estimated departure or arrival time, which will subsequently be used as the input for the DPA. Then, the DPA determines which flight delays will be propagated to the next time interval, thus affecting the demand of the next time period. With a new demand profile constructed, a new iteration begins. This continues until no more schedule updates are necessary.

QE

The QE is a numerical queueing model which is applied to each airport. The number of flights scheduled to take off and of land in each time period serves as the demand rate of the queue, and the capacity of runways serves as the service rate. Moreover, the order of accessing the queue is First Come First Serve (FCFS). In particular, the underlying queue is modelled as a $M(t)/E_k(t)/1$ system, with arrivals modelled as a non-stationary Poisson process and Erlang order k as the service time distribution. Since the exact computation for Erlang queues is time-consuming, an approximation method is used, which makes the AND highly efficient for computing local flight delays within each airport.

The QE approximates the evolution of the probability distribution for the number of flights waiting in the queue as the time proceeds. This facilitates the computation of quantities that measure performance, such as time-averaged daily local delay, expected upstream delay, etc.

DPA

The DPA captures delay propagation effects throughout the network of airports by determining which flights will propagate their delays so that they will be operated in an interval different from the one they were originally scheduled in. These flights are labeled as unprocessed flights[7]. For each time interval of a day, the DPA first computes a critical time t* which is set equal to the earliest updated departure or arrival time of unprocessed flights. The flights before this t* will be labelled as processed flights and do not propagate their estimated delays in the following time interval, thus not affecting demand for the next time interval. In contrast, the flights operated after this t* will be labelled as unprocessed flights and will cause a change to the demand of the next time interval, thus updating the demand profile. As a

[7]The time of a flight operation (landing or takeoff) is given by the final update of the time when this operation will take place.

result, new demand profiles for all the airports are obtained and will be used as the input for QE in the next iteration of the model.

The AND can be regarded as a very flexible framework, rather than a model. The QE can be implemented in different ways, using not only Erlang service times and a stochastic model but also deterministic demand and service processes. In addition, the number of airports can be increased beyond the 34 U.S. major airports which are covered in the current version of the AND model. Furthermore, since we compute the probability distribution of the number of aircraft in the queue over the day, we can design a computational module to calculate any statistic we may be interested in.

Although the AND has many advantages, it presents an estimation bias when we compare its delay estimates to the actual delay computed through the ASPM database. One reason for this bias is that the AND model does not consider the initial delay for each aircraft on any given day. For example, a flight may be scheduled to depart from Mexico City at 7:55am and arrive in Los Angeles at 10:30 am. But suppose that due to some mechanical problems, this aircraft cannot take off until 9:55 am. As a result, this aircraft would almost surely have a delay of 2 hours when it arrives at Los Angeles. Since the AND only takes the planned flight schedule as its input, this effect cannot be captured.

Moreover, when studying the flight delays within the NAS, the AND currently only has capacity profiles for 34 major airports in the U.S. For other airports, the AND assigns zero delay to any flight through them. This may lead to significant underestimation of delays for the entire network.

1.3.2 Passenger Delay Model

Bratu[Bra04] proposed the Passenger Delay Calculator (PDC) to compute passenger delays based on actual flight information, and proprietary booking data from an air-

line. The PDC can take into account missed connections and flight cancellations, if it has the actual flight information, such as whether a flight was cancelled or not, and whether the connection time between two consecutive flights is enough. In addition, the PDC includes a module that reallocates disrupted passengers by setting up a disrupted-passenger queue and also a maximum waiting time for any disrupted passengers, such as 12 hours in the daytime and 24 hours at night time. Following Bratu's work, Barnhart et al.[BFV10] generalized the PDC to 20 U.S. air carriers and introduced a passenger itinerary allocation model based on a discrete choice model.

The itinerary allocation model is used to allocate passengers to all the possible itineraries over a year. First, the model will extract attributes of itineraries, such as the local time of departure for the itinerary, the minimum seating capacity for the flights in the itinerary, etc. Then, a multinomial logit model is estimated by using the booking data from the sponsor airline. Using the logit model and macroscopic demand data, such as monthly itinerary demand and quarterly flight leg demand, each passenger can be allocated to a possible itinerary. This constructed passenger schedule information is used as the input for the PDC algorithm.

The PDC algorithm is executed in several steps. The first step is to determine disrupted passengers. Passengers who take two consecutive flights can be labelled as disrupted when either one of the two flights is cancelled or there is not enough time between these two flights for passengers to catch the next flight because of an arrival delay incurred by the first flight. In the second step, the disrupted passengers are ordered by their disrupted time in a recovery queue that operates according to some standard airline recovery policy. For the third step, the disrupted passengers will be reallocated to the earliest flight that has available seats and is leaving for their final destinations. During this assignment process, if the maximum waiting time threshold is triggered and the passenger still cannot be reallocated, the passenger is given a default delay equal to the maximum waiting time specified by the model's users. Eventually, delays for groups of passengers that have the same itineraries will be com-

puted. Based on these we can further compute various statistics for passenger delays.

The PDC is a powerful tool for computing passenger delays when actual flight information is available. But it requires major enhancement if it is to be used for computing passenger delays for hypothetical scenarios. In particular, we are interested in what impact an airline's flight cancellation strategy will have on passenger delays throughout an airline's network.

1.4 Research goal

The research presented in this thesis will propose an integrated model for testing the impact of different cancellation strategies on passenger delays for hypothetical scenarios. Since the AND outputs the estimated departure and arrival information for each flight, we will incorporate the AND into the PDC to compute the updated flight arrival and departure information after determining the flights cancelled. In this way, we can compute passenger delays for any hypothetical scenario. To this end, we will introduce a framework that integrates a Tail Recovery Model, Flight Cancellation Algorithms, a Refined Airport Network Delay (RAND) model, and the PDC at a daily scale. We will implement different flight cancellation algorithms and test their impact on the passenger delays.

We will also further explore the possibility to extend the integrated model to a monthly scale, or a scale that contains any length of time. In particular, we will extend the RAND to a monthly scale by constructing probability distributions for airport capacities through some classical machine learning algorithms, and validate our model by using the ASPM actual delay data.

1.5 Thesis outline

The main body of this thesis consists of two chapters: In Chapter 2, we provide a detailed description of the integrated model. Specifically, we describe multiple com-

ponents of the integrated model, including the Tail Recovery Model, the Capacity Profile Construction Model and the Refined Airport Delays Model (RAND), which is modified from the AND model proposed by [Pyr12]. Then we discuss multiple cancellation algorithms and evaluate their impact on passenger delays in terms of six criteria and seven typical days in 2007. We also compute passenger delay metrics on August 9th given different numbers of cancelled flights.

In Chapter 3, we extend the RAND model to compute flight delays on a monthly scale using different capacity profiles as input. These capacity profiles can be directly obtained from Federal Aviation Administration (FAA) reports or constructed by using classical machine learning algorithms on airport-level data. We validate our estimation of flight delays by using data of January 2008.

Chapter 2

The Integrated Model

Passenger delays have a high economic cost to the passengers themselves, to the airlines and to the economy at large. Therefore, it is important to evaluate the factors that influence passenger delays, such as different cancellation strategies and different weather conditions. However, the Passenger Delay Calculator (PDC)[BFV10] relies on an input of actually realized flight schedules to compute passenger delays. Thus, the PDC in its existing form cannot be used to calculate how the passenger delays would change in response to different cancellation strategies under different hypothetical scenarios. In this research, our goal is to use the Airport Network Delays (AND)[Pyr12] model as the tool to construct expected flight schedules and using this information to estimate passenger delays.

To this end, in this chapter, we will introduce an integrated model that incorporates the AND and the PDC at a daily scale, while the next chapter considers this combination at a monthly scale, or any other time scale. We differentiate models according to different time scales because they have different applications depending on their time scale.

The daily integrated model aims to study the impact of factors such as weather conditions or more explicitly airport capacities and the number of scheduled flights. In other words, this daily integrated model is only used for analyzing and estimating delays rather than predicting delays. The inability to predict delays can be explained by the large fluctuations of the number of flights operated by airlines on any specific

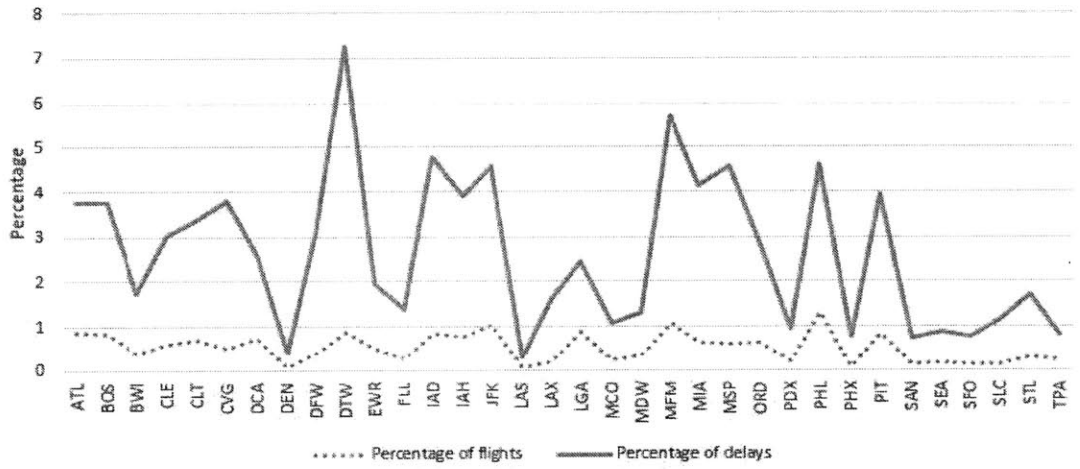
day and the unpredictability of weather conditions. For these reasons, we cannot expect too much for performing accurate flight delays or even passenger delays estimation. As a result, in the daily integrated model, we suppose that the capacity of each airport is known as well as the initial flight [1] delays for early flight of a day. Moreover, instead of studying a calendar day, we follow AND's setting that takes the study window from 3 am of a specific day to 3 am of the following day and denotes this time interval a "research day". AND takes 3 am as the starting time point of a day because it assumes no significant interplay from one day to the next so that (arrival) delays can be studied for an individual day. In Figure 2-1, we compute average arrival delays for late night flights, that is, flights scheduled to arrive before 3 am but actually arriving after 3 am. We observe that both the percentage of late night flights and percentage of late night flight delays are less than 1 %, which account for a very small portion of total number of flights and flight delays in a day. Hence, the inability of AND to handle late night flights will not have a significant impact on the flight delay estimation. When using the scheduled flight data as an input, we just simply ignore these late night flights. Similarly, we also ignore these flights for our daily RAND model which will be introduced later. However, for passenger delay computation, the no-interplay assumption seems to be problematic. Barnhart, Fearing and Vaze [BFV10] showed that the percentage of disrupted passengers receiving default delays is 33.5% for 6 hour / 12 hour delay limits (the maximum time any delayed passengers will wait) and 8.0% for 24 hour / 24 hour delay limits[2], while in the meantime the percentage of Passenger Disrupted is 3.3%. Although in terms of the number of disrupted passengers receiving default is roughly 0.5% to 1%, the total passenger delays caused by these passengers are significant.

Because of these cross-day effects, the daily integrated model can not grasp the overall mechanism of flight delays and passenger delays. In addition, we are motivated to extend the integrated model to predict delays at a macroscopic scale over an extended period of time. Therefore, in Chapter 3, we shall extend the daily integrated model to

[1] For definition of initial delays, see section 2.5

[2] the number before and after slash sign means the delay limits for day and night, respectively.

Figure 2-1: Number and percentage of arrivals delayed beyond the 24-hour study window (3 am to 3 am of the next day) (2007)



a monthly aggregated model with monthly flight schedules as input. This reduces the requirement of working with actual airport capacities and with actual initial delays for any given day and allows us to develop statistical estimates of delays based on long-term airport capacity data.

In section 2.1, we describe required input data sets for our integrated model as well as difficulties in combining the AND and the PDC due to the different data sets that these two models utilize. Section 2.2 presents the framework of the integrated model and the relationships among all of its components. In the remaining sections, we will introduce each component of the integrated model in detail. Moreover, we will present estimation results for both flight delays and passenger delays.

2.1 Data preparation

The AND model uses the Aviation System Performance Metrics (ASPM) database as a major input and so does the integrated model proposed in this chapter. The ASPM is compiled by the Federal Aviation Administration (FAA) and provides detailed data on flights to or from the 77 ASPM airports[3]. In addition to flights between the 77

[3]See Table A.1 and Table A.2

ASPM airports, the ASPM database also contains flights between any ASPM airport and an international airport or a non-ASPM airport in the U.S. In 2007, there was a total of 11,719,016 flights in the ASPM database. Among them[4], flights from or to the 34 major airports[5] account for 89.0% of total flights, while for 2013, this number is 90.7%. Figure 2-2 shows the number of flights arriving or departing from each of the 34 major airports. ATL and ORD serve the greatest number. In Figure 2-3, we show the number of flights in the database for some major airlines[6] in 2007 and 2013 (19 for 2007 and 16 for 2013), which represent 67.2% and 76.9% of total flights in the ASPM for 2007 and 2013, respectively. According to this figure, Southwest Airlines (SWA) and American Airlines (AAL) have the largest number of flights in 2007. However, in 2013 ExpressJet (ASQ) had the largest number of flights. This is because ExpressJet operates a very large number of regional flights[7]. In Figure 2-4, we consider flights only between 34 major airports, where this time the number of flights of ExpressJet is substantially lower than either 2007 or 2013.

The ASPM database combines several different sources of data, such as the Traffic Flow Management System (TFMS) data, the Airline Service Quality Performance (ASQP) data, etc. In particular, in 2007, the ASQP Data contributed 61.8% of the total number of flights in the ASPM.

The ASPM database consists of several modules. Two of these which are used in this research are the Airport Analysis module and the Individual Flights module. The data set of the Individual Flights module documents individual flights information, including the airline operating the flight, the tail number of the aircraft, the departure and arrival airports, the scheduled and realized arrival and departure time, etc. However, cancelled flights, military flights, and General Aviation flights are not recorded. The Individual Flights module facilitates the computation of flight delays for different airlines, different airports, etc. Figure 1-3 showed the average arrival

[4]We compute this statistic from ASPMs Individual Flights Module, which will be described later in this chapter.

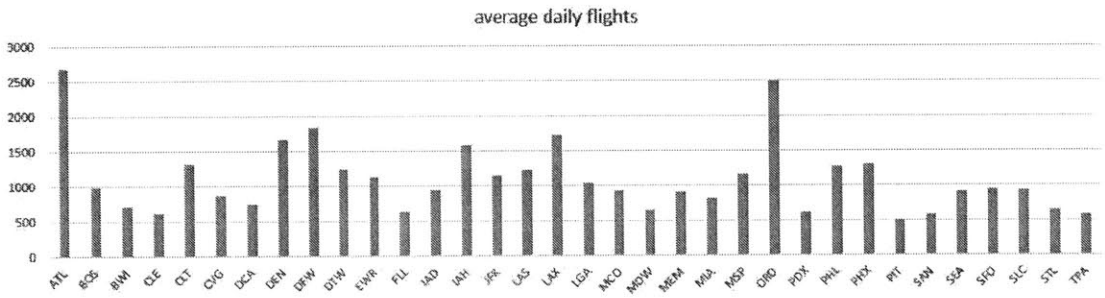
[5]See Table A.1

[6]See Table A.3

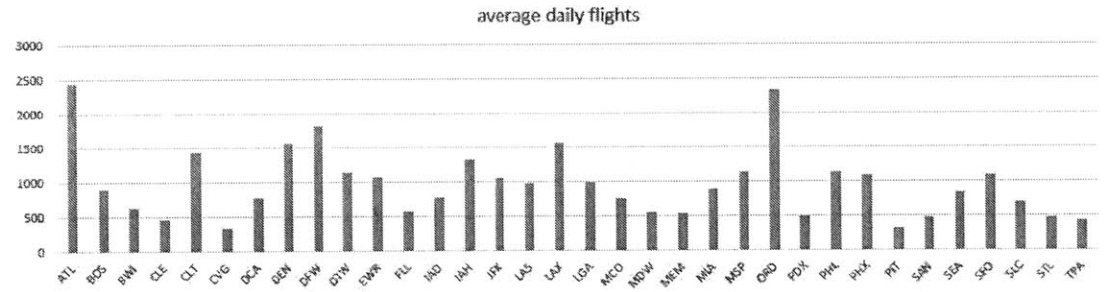
[7]The difference in the number of flights and number of airlines in 2007 and 2013 is due to airline mergers between 2007 and 2013. For example, Northwest Airlines (NWA) was merged into Delta Airlines (DAL) in 2008; Comair was merged into Delta in 2012.

Figure 2-2: Number of flights at 34 major airports in the ASPM

(a) 2007



(b) 2013

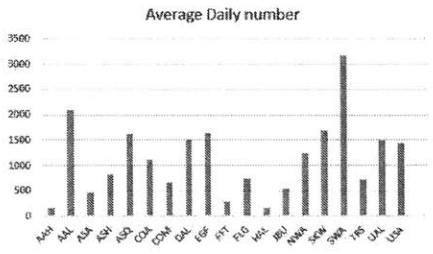


delays for 34 major airports in the U.S for 2007 and 2013, while Figure 2-5 shows average arrival delays for 20 major airlines for 2007. We have used the Individual Flights module to obtain scheduled flight time information, which is an important input for the Refined Airport Network Delays (RAND) model, while the actual flight time information will serve to validate the flight delays estimation obtained by the RAND. However, the original data set cannot be used directly as an input for the RAND, since a significant percentage of tail numbers of flights is artificial[8]. For example, in 2007, 38.2% of flights lack actual tail numbers. This situation leads to a potential underestimation of flight delays in that the RAND models the flight delays via the itinerary of each aircraft, but the propagation effect of delays through the same aircraft will not be captured if some flights which in reality are operated by the same aircraft, are assumed to be operated by different aircraft in the RAND model.

[8]In the ASPM, tail numbers starting with Capital letter N are actual tail numbers. Other tail numbers just simply follow their flight numbers, concatenating with airlines' ICAO code in the beginning. For example, "N626AS" is an actual tail number, whereas "AAH923" is an artificial tail number which associates to a flight from "AAH" airline and whose flight number is 923.

Figure 2-3: Number of flights of major airlines in the ASPM

(a) 2007



(b) 2013

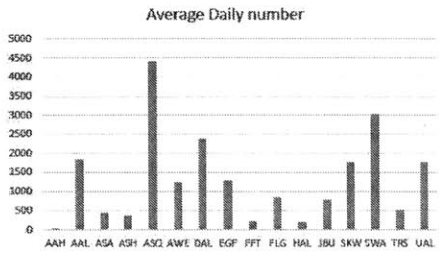
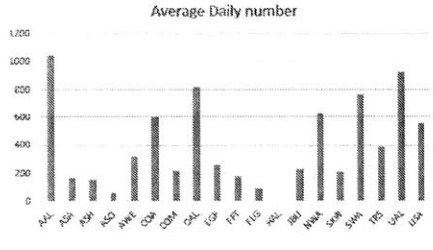
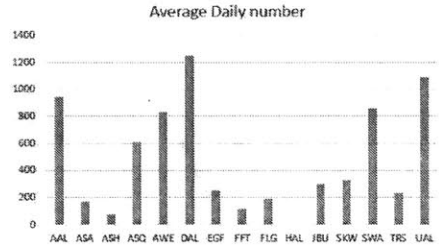


Figure 2-4: Number of flights of major airlines between the 34 airports in the ASPM

(a) 2007



(b) 2013

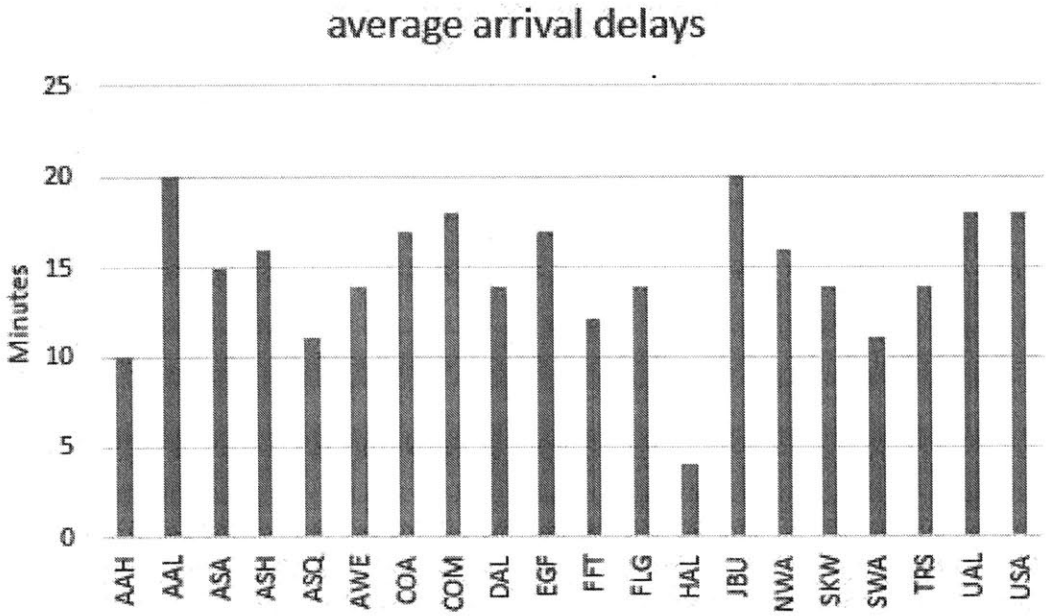


Therefore, we introduce a tail recovery model to associate flights with artificial tail numbers to those with actual tail numbers (see Section 2.3).

The data set of the Airport Analysis Module documents weather conditions hourly for each of 77 ASPM airports, as well as tens of other metrics, such as estimated departure and arrival rates, average gate arrival delays, etc. In this thesis, we will study the impact of weather conditions on capacities and then delays. In particular, we will construct a daily capacity profile for each airport based on meteorological conditions and further use the capacity profiles as an input for the RAND (See Section 2.4). In Chapter 3, we will introduce a Markov Chain Model to dynamically construct the meteorological conditions and then capacity profiles, based on which flight delays for a period of time, such as a month or a year, can be computed.

The second major database on which we shall rely was developed at MIT by combining a number of different databases of the FAA and the U.S. Bureau of Transportation Statistics(BTS). We denote it by HDFB (Hybrid database combining databases from

Figure 2-5: Average Arrival Delays for major airlines in 2007



FAA and BTS). This is the major input database for PDC [BFV10]. In our integrated model, we will continue using this database as the input to the PDC. The HDFB combines the Airline Service Quality Performance (ASQP) database, the Schedule B-43 Aircraft Inventory (B-43), T-100 Domestic Segment (T-100) and the Airline Origin and Destination Survey (DB1B) maintained by the BTS, as well as the Enhanced Traffic Management System (ETMS) database maintained by the FAA. The ASQP database provides scheduled flight information similar to that of the Individual Flight Module of the ASPM, such as scheduled and actual arrival and departure time, arrival and departure airports, tail numbers, etc. Moreover, it also contains cancelled flights information. The B-43 database provides seat capacity information for most airlines, but can only match 75% of flights in the ASQP due to the artificial tail number problem in the ASQP; the ETMS provides International Civil Aviation Organization (ICAO) aircraft equipment code and helps to match many of the remaining 25% flights in the ASQP by using seat capacity; the T-100 and the DB1B provide aggregate passenger demand.

There is, however, a big difference between the ASPM and the HDFB in terms of the

number of flights they contain. The total number of flights in the HDFB in 2007 is only 7,455,428, compared to 11,719,016 in the ASPM. This difference is due to the fact that the ASQP is only one of the data sources that provide flight information for the ASPM. Moreover, the ASPM does not contain cancelled flights, whereas the original ASQP database does. This difference leads to a problem when combining the PDC and the AND models, since we cannot use a unified database as an input for the integrated model. The RAND requires the full set of flights as an input, or flight delays will be underestimated. However, the PDC cannot use the full set of flights as an input, since many flights do not have seat capacity information and belong to other airlines beyond the 20 covered by the PDC. Therefore, at the cost of losing a substantial number of flights, which are not reported in the HDFB, we use the Individual Flight Module of the ASPM as an input for the RAND, but match the result from RAND to the HDFB before computing passenger delays.

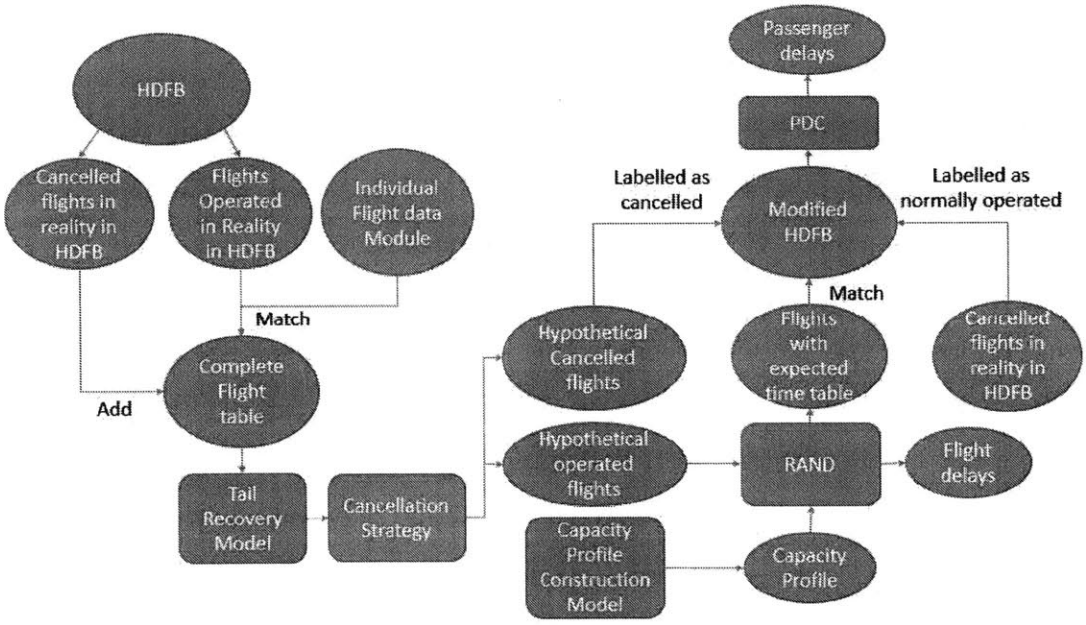
2.2 Overview of the Integrated Model

The integrated model provides a framework for quantifying aggregate passenger delays at a daily scale based on the flight schedule information, flight cancellation information and airport conditions (weather and capacity conditions). It turns out to be very flexible, since we can implement different cancellation and operating strategies, either existing or hypothetical, according to decisions made or suggestions given by airline managers, airport managers and officials from the FAA, and observe how the entire airport network delays vary in response to these strategies.

The framework of the integrated model at a daily scale is illustrated in Figure 2-6, where squares represent processing steps, and ellipses represent data sets.

Before using this integrated model, we prepare three data sets for a specific day from the HDFB, the Individual Flights Module and the Airport Analysis Module of the ASPM. In the beginning stage of the framework, we combine the HDFB data set and the ASPM flight data set. In particular, the HDFB data set is divided into two

Figure 2-6: The framework of the integrated model



parts, flights that were actually operated and flights that were cancelled in reality. We match the former to the ASPM flight data set in order to obtain an integrated data set since both sets contain only actually performed flights. But in order to implement cancellation strategies in hypothetical scenarios, we also assume that all the actually cancelled flights were not cancelled and include them into the integrated data set. Since a significant percentage of flights in this integrated data set have artificial tail numbers, we employ a tail recovery model to recover default tail numbers. In particular, flight-specific information will be processed to construct entire itineraries for each individual aircraft.

After constructing in this way a complete set of flights, a specific cancellation strategy is executed to generate a specified number of hypothetically cancelled flights. The remaining not cancelled flights are fed into the RAND along with capacity profiles for all the 34 airports. The capacity profiles are either constructed according to the FAA Benchmark report (2004)[ACB04] or based on construction methods described in Section 2.4 and in Chapter 3. The RAND estimates flight delays for each flight and also outputs expected arrival and departure time information for each flight. Based

on this flight data set (note that it also identifies hypothetical cancelled flights), the next step is to modify the PDC input dataset – HDFB. In particular, we treat flights which were cancelled in reality as normally operated flights and flights that are hypothetically cancelled as actually cancelled flights in HDFB. Moreover, we update the actual arrival and departure information by replacing it with expected arrival and departure information computed through RAND. Eventually, we use this modified HDFB dataset as an input to the PDC and compute passenger delays.

The matching of data sets is executed twice in this framework. It first takes place between flights actually operated in the HDFB dataset and in the ASPM dataset before executing the Tail Recovery Model (see Figure 2-6). The second match takes place between flights in the table of the expected arrival and departure times and the HDFB dataset before producing modified HDFB dataset for the PDC (see Figure 2-6). The number of flights that can be matched at this stage determines how many flights will be updated in the HDFB in terms of their arrival and departure time, thus directly affecting the accuracy of the estimation of passenger delays. In Table 2.1, we use a random cancellation strategy[9] as an example and show the extent to which we can match the HDFB and the flight module data set of the ASPM, including cancelled flights. For the seven days shown in the table, the percentage of matched flights is always well above 90%.

[9]See section 2.7

Table 2.1: Performance of matching the HDFB dataset and the ASPM dataset for the random cancellation strategy for seven days in 2007

Date	Nov 23rd	Jun 14th	Jun 20th	Aug 9th	Mar 2nd	Feb 24th	Feb 14th
Flights Actually Cancelled	52	284	515	819	1238	2121	4385
Flights from the ASPM	26947	35775	35522	34084	34225	27604	31544
Flights from the HDFB	16946	21663	21639	21728	21305	17988	20458
Matched Flights	16166	20979	20254	20461	20836	17484	19742
Percentage of HDFB flights matched	(95.4%)	(96.8%)	(93.6%)	(94.2%)	(97.8%)	(97.2%)	(96.5%)

From the leftmost column to the rightmost column, the number of cancelled flights increases. We will keep using this order for these seven days in all subsequent tables.

In section 2.3 through section 2.8, we will present each component of the integrated model in detail. We select seven days (November 23, June 14, June 20, August 9, March 2, February 24 and February 14) in 2007 as computational examples and show results for these days for each module in the integrated model. These selected days have different numbers of actually cancelled flights, with Feb 14 and Nov 23 being days with the most and the least number of flights cancelled in 2007, respectively. In particular, the largest number of cancellations is due to the Valentine's Day storm, of February 13-14, 2007, while November 23 is the Friday after Thanksgiving. These days are representative in the sense of having a wide range of delays and cancellations. The basic profile of these days including the number of flights actually cancelled and the total number of passengers from HDFB is presented in Table 2.2.

Table 2.2: The basic profile for 7 selected days

Date	Nov 23rd	Jun 14th	Jun 20th	Aug 9th	Mar 2th	Feb 24th	Feb 14th
Flights cancelled	52	284	515	819	1238	2121	4385
Number of passengers	1186717	1545767	1461023	1507411	1489961	1036570	1141024

2.3 The Tail Recovery Model

The tail recovery model recovers artificial tail numbers for the integrated data set that merges the HDFB and the flight data set in the ASPM, thus providing full aircraft itineraries. This is very important for modelling the delay propagation effect in the airport network in the RAND. In the integrated data set, roughly 30% to 40% of flights do not have actual tail numbers. For instance, on Aug 9th, the total number of flights is 34084, of which 35.5% did not have actual tail numbers. Moreover, there is a small percentage of flights lacking aircraft type information. For example, this number is 224 on August 9th. The tail recovery model has a significant impact on the flight delays estimation. For instance, following the example of Chapter 1, suppose, for a specific day, an aircraft departs from Chicago O'Hare Airport (ORD), where serious congestion occurs and the delay time is about 30 minutes, to Los Angeles (LAX) via Phoenix (PHX), where little congestion occurs and the delay time is about 5 minutes. Under this setting, if the tail number is correct for this aircraft, or in other words, two flight legs share the same tail number, AND computes the arrival delay at LAX 30 minutes and the arrival delay at PHX 35 minutes, leading to a total arrival delay 65 minutes for this aircraft. On the other hand, if the tail number for two flight legs are different, these two flights legs are treated as if they were operated by two different aircraft. Aircraft A makes the first flight, and aircraft B the second. Assuming all the other flight schedules are the same as in the previous scenario, Aircraft A still suffers 35 minutes of delay computed by the AND, whereas Aircraft B takes off on time and does not suffer delay at all! As a result, the total arrival delay is only 35

minutes, as opposed to the correct 65. Therefore, it is important to recover the tail numbers as accurately as possible.

Our tail recovery algorithm is shown in Algorithm 1. Basically, it infers the tail numbers of flights without actual tail numbers by relying on flights with existing tail numbers and making additional assumptions, such as that matched flights should belong to the same airline, have the same aircraft types, and match scheduled time information.

Algorithm 1 The Tail Recovery Algorithm

```
1: MinTT ← Default Minimum transfer time
2: MaxTT ← Default Maximum transfer time
3: N ← the number of flights
4: F ← the list of all the flights including different attributes, such as tail numbers
   (TN), aircraft types (AT), airlines (A), scheduled arrival time (SAT), scheduled
   departure time (SDT), original airport (OA), destination airport(DA), actual ar-
   rival time (AAT), actual departure time (ADT)
5: N_d1 ← the number of different tail numbers
6: N_d2 ← 0
7: while N_d1 ≠ N_d2 do
8:     Sort flights by A, TN
9:     for i ← 1 to n do
10:        if f[i].AT = ∅ and f[i].TN ≠ ∅ then
11:            j ← i + 1
12:            if f[i].A = f[j].A then
13:                if f[i].TN = f[j].TN and f[j].AT ≠ ∅ then
14:                    f[i].AT ← f[j].AT
15:                    Continue
16:                end if
17:            end if
18:            j ← i − 1
19:            if f[i].A = f[j].A then
20:                if f[i].TN = f[j].TN and f[j].AT ≠ ∅ then
21:                    f[i].AT ← f[j].AT
22:                    Continue
23:                end if
24:            end if
25:        end if
26:    end for
```

```
27:     Sort flights by A, AT, SDT
28:     for i ← 1 to n do
29:         if f[i].TN = ∅ then                          ▷ f is an abbreviation for flight
30:             j ← i + 1
31:             while f[i].A = f[j].A, f[i].AT = f[j].AT do
32:                 if f[i].DA = f[j].OA, f[i].SAT + MinTT ≤ f[j].SDT, f[i].SAT +
    MaxIT ≥ f[j].SDT, f[i].AAT + MinTT ≤ f[j].ADT, f[i].AAT + MaxIT ≥
    f[j].ADT and f[j].TN ≠ ∅ then
33:                     f[i].TN ← f[j].TN
34:                     break
35:                 end if
36:                 j ← j + 1
37:             end while
38:             j ← i − 1
39:             while f[i].A = f[j].A, f[i].AT = f[j].AT do
40:                 if f[i].OA = f[j].DA, f[j].SAT + MinTT ≤ f[i].SDT, f[j].SAT +
    MaxIT ≥ f[i].SDT , f[j].AAT + MinTT ≤ f[i].ADT, f[j].AAT + MaxIT ≥
    f[i].ADT and f[j].TN ≠ ∅ then
41:                     f[i].TN ← f[j].TN
42:                     break
43:                 end if
44:                 j ← j − 1
45:             end while
46:         end if
47:         N_d2 ← the number of different tail numbers
48:         N_d1 ← N_d2
49:     end for
50: end while
```

The algorithm is executed iteratively in two phases, recovering aircraft types and tail numbers, until the number of aircraft stays the same for two consecutive iterations.

For each iteration, in Phase 1 (Line 8 - Line 26), the algorithm first sorts flights according to airline and tail number. It then tries to recover aircraft types for each target flight, that is, the flights without aircraft type information but with tail numbers. Since we have sorted all the flights, for each target flight, we only need to look at the previous and the next flight in the list. If one of them matches the target flight's tail number and also has aircraft type information, the aircraft type of the target flight is also known. In Phase 2 (Line 28 - Line 49), we first sort flights according to airline, aircraft type and scheduled departure time. Then, for each target flight (flights with incorrect tail numbers but having aircraft type information), the algorithm searches flights that are before and after the target flight in the list. If a flight has the same airline and aircraft type as the target flight, its destination airport matches the arrival airport of the target flight, and has a reasonable time interval between arrival (both scheduled and actual) time and the departure time (both scheduled and actual) of the target flight and itself, these two flights are recognized to be operated by the same aircraft, thus the tail number of the target flight can be recovered. Since actual or default tail numbers provide useful information for recovering aircraft types, the algorithm feeds these actual or default tail numbers into Phase 1, thus starting the next iteration. After several iterations (in practice, the typical number is 5 to 10), the algorithm will terminate, when no more tail numbers can be recovered by continuing to iterate.

Phase 1 is a little different from Phase 2 in terms of the scope of searching for matched flights for each target flight. Phase 1 only considers the previous and next flight on the list, whereas Phase 2 considers more. The reason for this is the following: The number of flights with missing aircraft types is relatively small (about 1% of total flights), thus consecutive flights in the list with missing aircraft types are very rare. Also, owing to the precise match condition of tail numbers, the algorithm can always find either the previous or the next flight to recover the aircraft type of a target flight. However,

this will not work for recovering tail numbers. We only have imprecise matching conditions such as name of airline and aircraft type in Phase 2, therefore the arrival and departure time of flights are taken into account in the matching conditions. In particular, assuming other matching conditions are satisfied, if the difference of the arrival time of a flight and the departure time of another flight is greater than the minimum transfer threshold and less than the maximum transfer threshold, then we identify these two flights as matched.

Even though this algorithm recovers almost all the tail numbers for aircraft with artificial tail numbers, flights with both artificial tail numbers and missing aircraft types cannot be recovered. However, this number is very small (see Table 2.3) for unmatched flights.

In the tail recovery algorithm, minimum and maximum transfer times are important parameters. In practice, according to experiments, 30 minutes for minimum transfer time, and 720 minutes for maximum transfer time are best for the purpose of recovering most of the flights with artificial tail numbers.

In Table 2.3, we present the total number of flights for seven selected days (same as before) as well as the number of different tail numbers before and after application of the tail recovery model. The substantial decrease in the number of different tail numbers indicates the good performance of the tail recovery model. Moreover, the number of unmatched flights is also minimal compared to the total number of flights and can be safely ignored.

Table 2.3: Results for the tail recovery model

Date	Nov 23rd	Jun 14th	Jun 20th	Aug 9th	Mar 2th	Feb 24th	Feb 14th
Total flights in the ASPM	26947	35775	35522	34084	34225	27604	31544
Different tail numbers before recovery (percentage of total flights)	13335 49.5%	16703 46.7%	16789 47.3%	16738 49.1%	16533 48.3%	14940 54.1%	17432 55.3%
Different tail numbers after recovery (percentage of total flights)	9930 36.9%	12091 33.8%	12128 34.1%	12357 36.3%	12047 35.2%	11095 40.2%	12355 39.2%
unmatched flights (after both phases)	22	14	17	15	32	19	28

2.4 Capacity Profile Construction Model

The capacity of an airport describes the rates at which the airport can process arrivals and departures during each time interval of a day. In the AND (RAND) model, capacity profiles for 34 major airports are constructed for any given day and used as an input of the AND (RAND). When the Queueing Engine estimates local airport delays based on its underlying numerical queuing model, the capacity of an airport can be thought of as the number of flights that can be processed per unit of time by the queueing system.

In this thesis, we will use the following types of capacity profiles:

1. The FAA benchmark three-level capacity profile. This benchmark is given by Airport Capacity Benchmark Report 2004[ACB04]. As shown in Table A.1, this benchmark provides three levels of capacities for each of 34 airports as well as their corresponding occurrence probabilities. In particular, the benchmark provides an Optimum Capacity (OPT CAP), a Marginal Capacity (MARG CAP), and an Instrument-Flight-Rules (IFR) Capacity, which represent weather con-

ditions ranging from good to intermediate to low visibility. Thus, airport capacities vary from high to medium to low levels, respectively. From the probability associated with each level, we also observe that good weather conditions prevail on most days at all the airports. This type of capacity profile can be used for purposes of a macroscopic study, but is not suitable for tracing the capacity evolution during a specific day. However, in Chapter 3, when we will estimate flight delays on a monthly scale, this profile will be used as a benchmark and serve as an input for the monthly RAND model for estimating flight delays.

2. The hourly estimated capacity levels contained in the Airport Analysis Module of the ASPM database. This profile contains hourly capacities of airports each day. In this Chapter, we use this capacity profile as an input for the day-scale integrated model. In practice, this hourly capacity can be estimated from historical data by using any popular regression machine learning model.

3. Probabilistic capacity profiles constructed through capacity profile construction models that we have developed. This capacity profile has only two levels, "High" and "Low", which is a simplification of the original capacity profile in which capacity can vary continuously over a range of values. To produce a predictive model for this classification problem, we use some classical machine learning models, such as the logistic regression model, the Classification And Regression Trees (CART) model, etc. In Chapter 3, we will compare flight delays estimated based on capacity profiles using these algorithms to those based on the FAA benchmark capacity profile.

2.5 The Refined Airport Delays Model(RAND)

The Refined Airport Network Delays (RAND) model is a modified version of the Airport Network Delays (AND) model. The AND model captures the occurrence of local airport delays by using the Queueing Engine (QE), and quantifies the propagation of flight delays by employing the propagation algorithm. By integrating these two

components, the AND captures the mechanism of flight delay generation and propagation throughout the network. However, the AND model includes two assumptions that may lead to a substantial underestimation of the actual flight delays provided in the Individual Flight Module of the ASPM.

The first assumption is the zero-delay assumption at the other airports beyond the 34 major ones. The flights whose departure and arrival airports are both in these 34 airports account for 67.2% of the total flights in the ASPM. The remaining flights either have only one of their departure or arrival airports in the 34 major airports or neither their destination nor their origin airport is one of the 34 airports. The AND model only considers the 34 major airports in the U.S., treating all the other 43 airports in the ASPM database as a virtual airport, "ZZZ". Moreover, AND assumes no arrival or departure delay is incurred at "ZZZ" for any flight, which may not be in accordance with reality. For example, a flight may be scheduled to depart from Albuquerque (ABQ) at 12:00 and arrive at Dallas/Fort Worth (DFW) at 14:40 but in reality, its departure time and arrival time may have been 12:50 and 15:34, respectively (this is a real example from the ASPM database). The AND assumes no flight delay occurred at ABQ since ABQ does not belong to the 34 major airports, thus this 50 minutes delay is ignored. Using the ASPM Individual Flight Module and calculating the average delays at the other 43 airports, we show in Figure 2-7 and Figure 2-8 that the average arrival and departure delays for these 43 airports were significantly different from zero in 2007. In RAND, we will relax this zero-delay assumption of the AND by adding a default arrival delay and departure delay to each flight from or to these airports. These additional delays increase significantly the arrival and departure delays for airports in the network, and make them closer to the actual ones.

The second assumption is the zero-delay assumption for "initial flights". An initial flight refers to the first flight in an aircraft's itinerary on a specific day. As indicated earlier, the AND assumes a starting time of a day, for example, 3 am, and does not consider any delays caused by flight delays from the previous day. For example, on a

Figure 2-7: Average Arrival Delays for 43 other airports in 2007

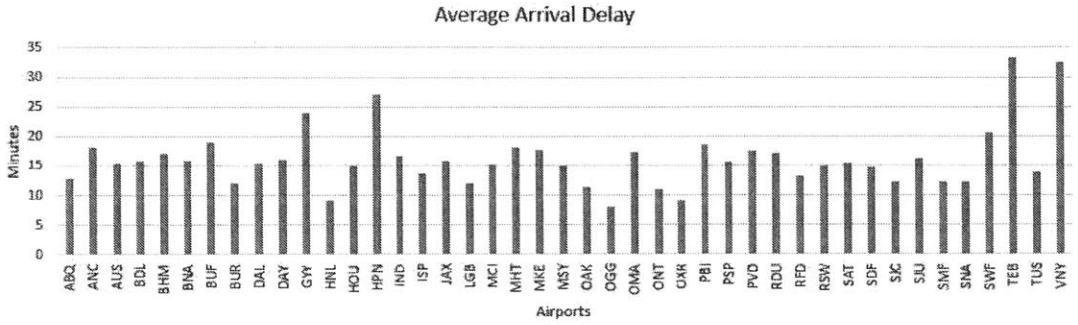
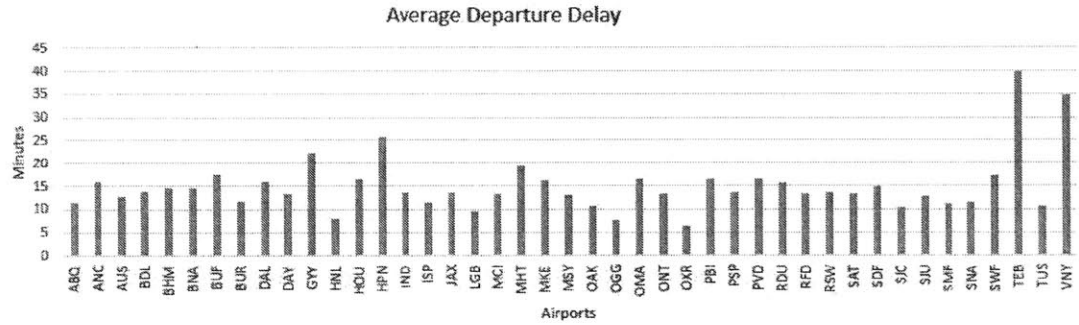


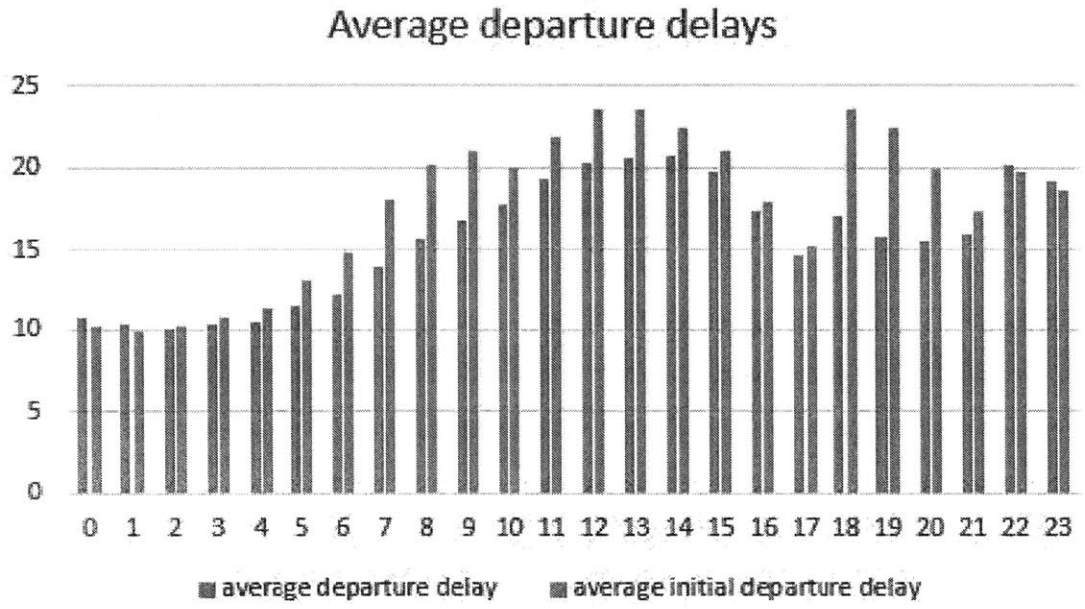
Figure 2-8: Average Departure Delays for 43 other airports in 2007



specific day, an aircraft may be scheduled to depart from LGA at 7:55 am and arrive at STL at 9:35am, followed by a flight from STL at 10:37 am to LGA at 2:00 pm. In this scenario, the flight from LGA to STL is the initial flight. The delay associated with the initial flight is called initial delay. When an initial flight is scheduled early in the day, its initial delay cannot always be explained by local congestion or a delay propagation effect from an upstream airport since the initial delay is generated in the early hours. However, the AND model ignores these initial delays and assumes the initial departure and arrival delays for each airport to be zero, thus leading to inaccuracy in terms of computing delays.

To reduce the negative effect of this assumption, we take into account the initial delay that cannot be explained by local delay effects and delay propagations explored by the AND model. However, exactly identifying this initial delay effect turns out to be very difficult, since during the daytime the initial delays are probably caused either by mechanical problems of the aircraft themselves or by local congestion. To remain

Figure 2-9: Average (Initial) Departure Delays in 2007



on the conservative side, we only introduce initial delays to flights scheduled to arrive and depart during the early hours of a day. As shown in Figure 2-9, from 3 am to 6 am, average initial departure delays are small. Moreover, most flights during this period of time can be counted as initial flights. Therefore, we directly use the actual departure time as their departure time instead of using scheduled departure time as before. Likewise, we introduce initial flight arrival delays to flights scheduled to arrive during 3 am to 6 am but not scheduled to depart during 3 am to 6 am. For flights' arrival time later than 6 am, we just assume that the additional initial delay for any flight is zero. Some statistics show the magnitude of these initial delays. In 2007, 41.4 % of flights scheduled to arrive between 3 am and 6 am had positive-minute delays, while 43.1 % of flights scheduled to depart within these hours had positive-minute delays. Figure 2-10 and Figure 2-11 present the histograms of flights during these hours having departure and arrival delays, respectively. Surprisingly, the γ distribution fits these two histograms very well. We will further study these underlying distributions for the yearly initial departure and arrival delay data in Chapter 3, using the one-sample Kolmogorov-Smirnov test to show that we cannot reject the γ distribution

as a possible underlying distribution for these data. In Chapter 3, we will also feed the initial delays sampled from these underlying distributions to each day in a time period of interest in the extended RAND model, and estimate flight delays on average. However, the histograms of initial departure delays vary significantly from day to day, as shown in Figure 2-12. Therefore for the input to the daily integrated model in this chapter, we replace the scheduled departure and arrival time with actual time.

With initial departure and arrival delays added to flights in the early time of the

Figure 2-10: Histogram of initial departure delays for flights scheduled to arrive between 3 am and 6 am in 2007 (in minutes)

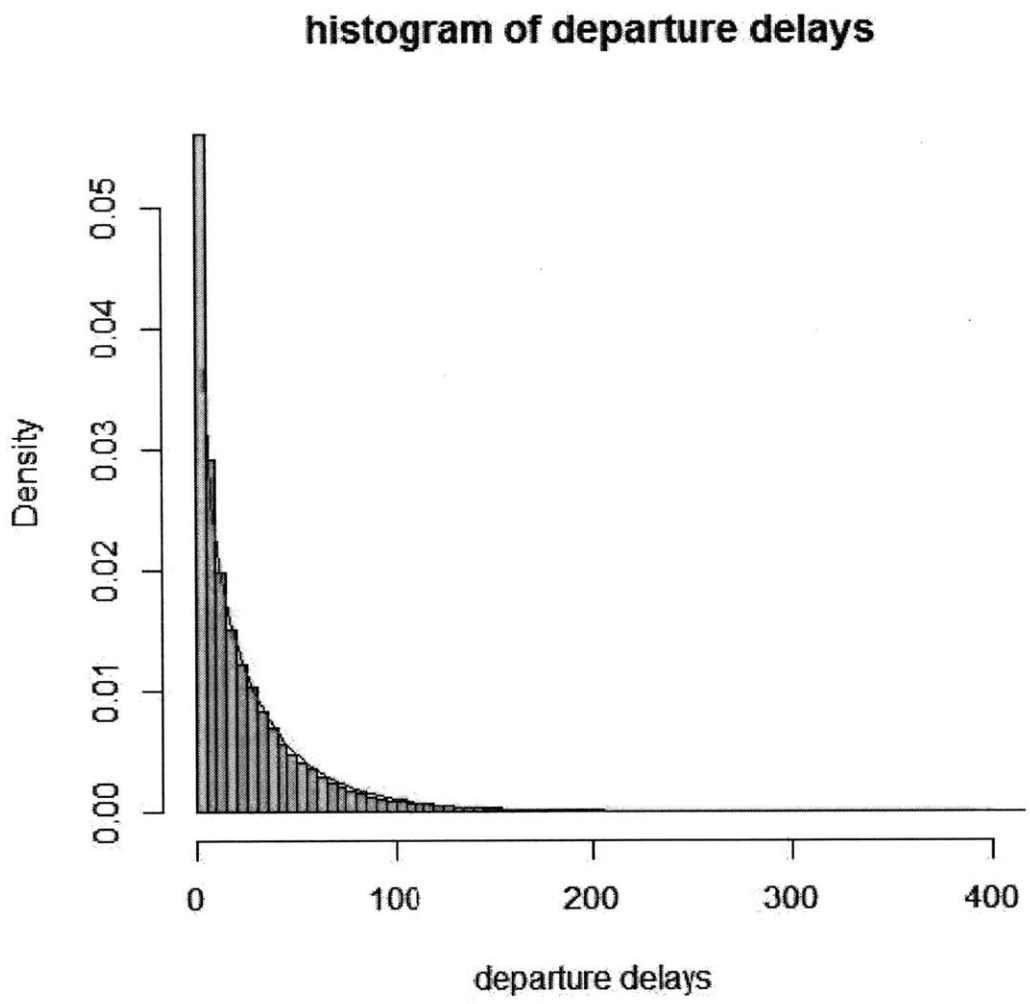


Figure 2-11: Histogram of initial arrival delays for flights scheduled to arrive between 3 am and 6 am in 2007 (in minutes)

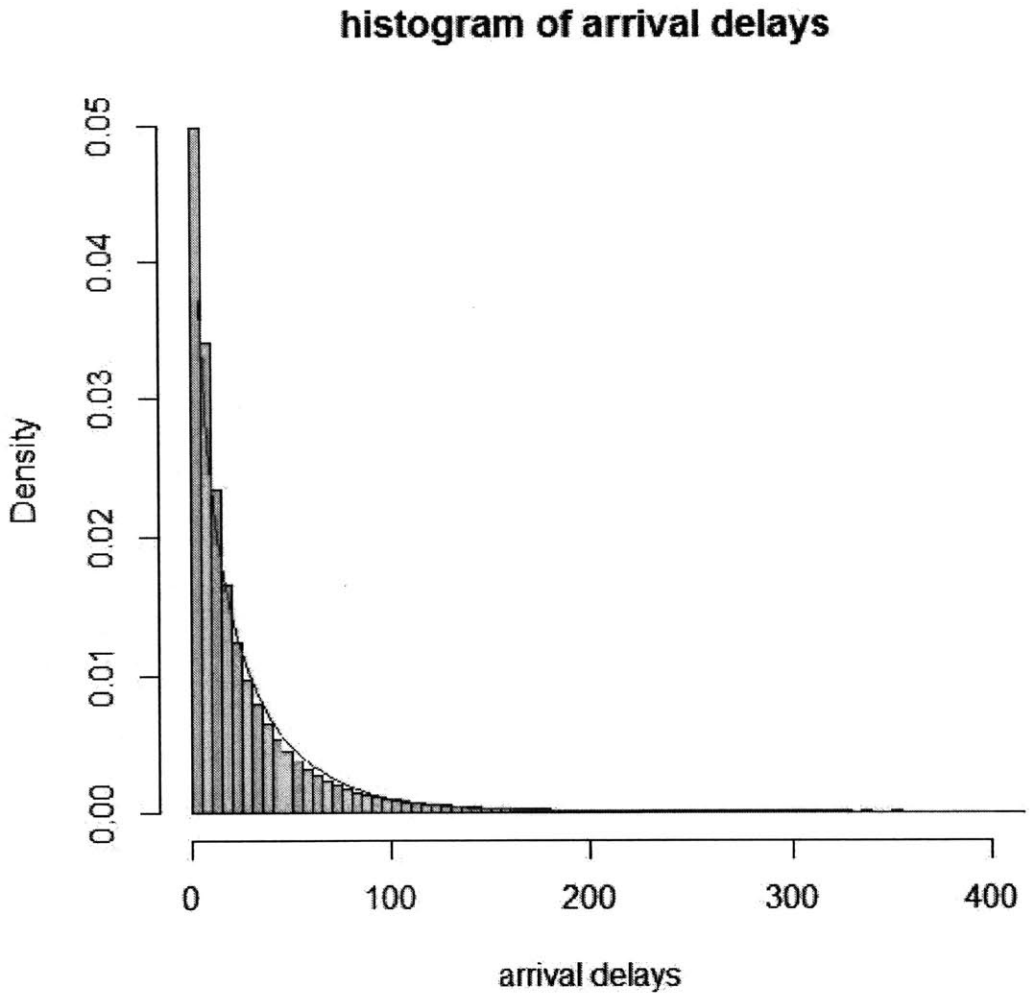
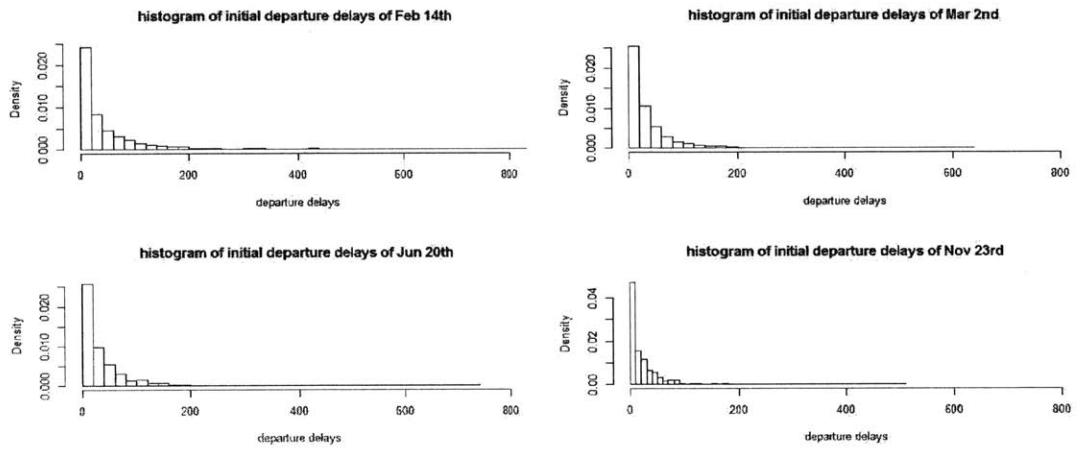


Figure 2-12: Histogram of initial departure delays for flights scheduled to arrive between 3 am and 6 am for four days with different magnitude of congestion in 2007 (in minutes)



day and default delays assigned to flights via 43 minor airports, the RAND model improves the flight delay estimates in the sense of approaching the delays calculated from the data in the ASPM flight level module. The estimation of arrival delays over the 34 major airports for the seven representative days of 2007 is shown in Figure 2-13, where a red bar represents the delays calculated from ASPM, while a blue one represents delays estimated by the RAND.

Figure 2-13: Estimation of arrival delays over the 34 major airports for the seven representative days of 2007

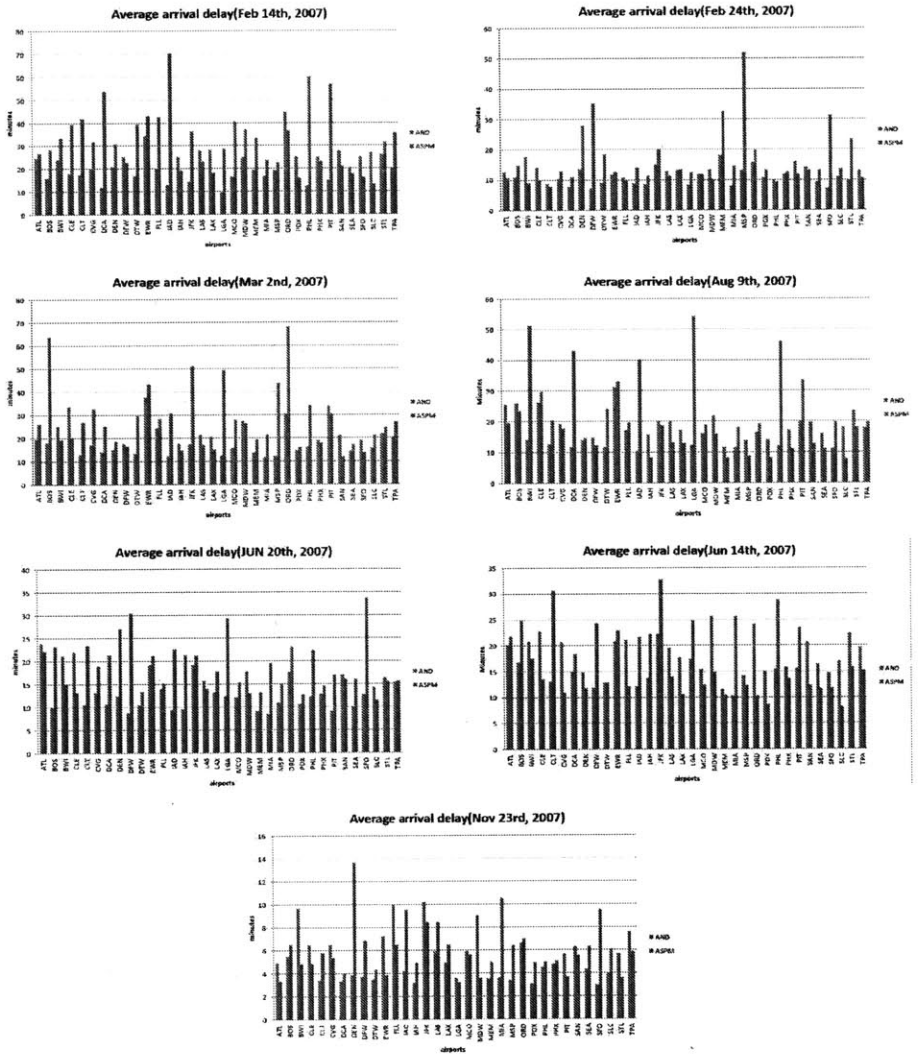


Table 2.4: Some statistics for average arrival delays estimation over the 34 major airports for the seven representative days of 2007 (delays are given in minutes)

Date	Nov 23rd	Jun 14th	Jun 20th	Aug 9th	Mar 2th	Feb 24th	Feb 14th
Flights Cancelled	52	284	515	819	1238	2121	4385
average arrival delays in ASPM	7.39	19.38	21.25	23.24	30.08	18.88	35.67
average arrival delays from RAND	5.09	15.47	12.99	18.28	16.33	10.76	20.9
number of airports with underestimated delay	20	14	26	17	23	22	22
number of airports with estimation error within 30 % of ASPM value	13	13	17	11	17	20	11

We summarize our observations from Figure 2-13 in the following list and also elaborate on some of them in Table 2.4:

1. Higher delays correspond to higher cancellation rates. With the exception of Feb 24, the average arrival delay increases monotonically with the cancellation rate.

2. For most airports, the average arrival delays over all 34 airports estimated by the RAND matches the actual arrival delays obtained from the ASPM reasonably well in terms of estimation error. In particular, the estimated average delay for more than one third of the 34 airports is within 30 % of the actual delays. Moreover, except for Feb 24, and Jun 20, more than 25 airports have an estimation error of less than 50 %[10].

3. Except for Jun 14 and Aug 9, RAND underestimates the arrival delays at

[10]This observation is not presented in Table 2.4.

roughly more than two thirds of airports.

4. For airports with moderate arrival delays (ex. MCO in the scenario of Feb 24), the estimation of RAND is quite close to the true delays given by ASPM.

5. For airports with extremely large arrival delays (ex. IAD for Feb 14th, 2007 and MSP for Feb 24th), the RAND is not able to give a good estimate.

2.6 Experiments with the RAND model

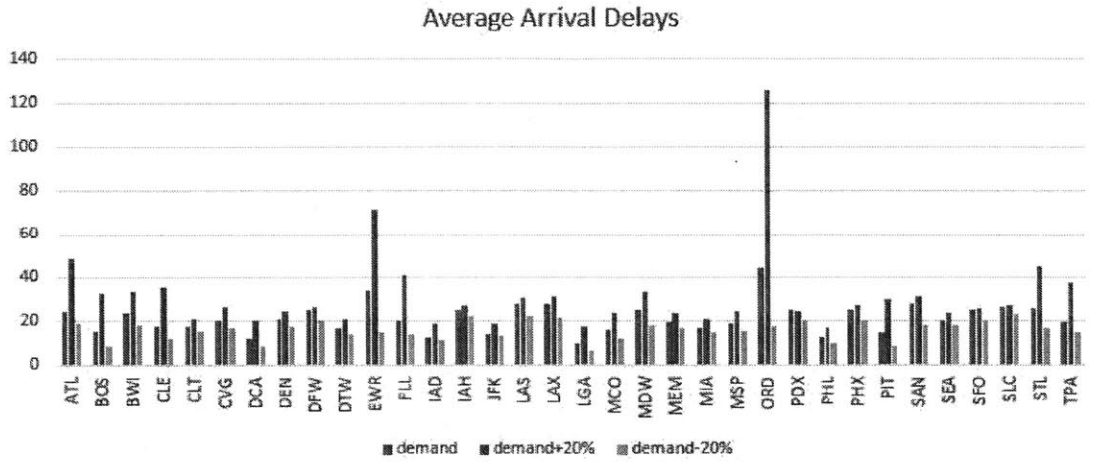
In Section 2.5, we described the RAND model, which is mainly comprised of the AND model, but additionally takes into consideration the initial delays for early flights of a day and the default delay for flights via the airports other than 34 major airports. The RAND model calculates flight delays mainly based on two factors, the number of flights scheduled for a day (flight demand) and the capacities of 34 major airports. Moreover, the flight delays are also affected by the minimum turn around time, the maximum in air saving time for all the flights, the default delay parameters, queueing model parameters such as the Erlang orders, etc. All these secondary inputs for the RAND model are assumed to be set in accordance with our prior knowledge of the airport system or directly from the available data. The minimum turnaround time is the minimum time during which the aircraft must remain at the gate. It varies over different airlines and different aircraft types and different airports. The maximum in air saving time is the time that a flight can make up by speeding up en route. This is set to be 10 minutes in the model. The Erlang order is chosen to be 9. However, this parameter has proven to be of limited importance to the flight delay estimation. With all these factors fixed, we test the sensitivity of RAND to the two principal inputs for a specific day, February 14th.

Experiment 1: We consider three scenarios having different levels of flight demand. The first one is exactly the actual scenario on February 14th, whose airport capacity profiles come directly from the Airport Module in the ASPM database. We construct the second scenario without changing the capacities of the airports, but increase the

number of arrival flights at each airport by 20%. In particular, this scenario can be achieved by randomly choosing 20% of existing flights for each airport, giving them tail numbers different from the existing ones, and including them in the additional flight demand. With the same method, we construct the third scenario under the same conditions as the second scenario except randomly cancelling 20% of arrival flights for each airport. We then use the RAND model to estimate the average arrival delays based on these three scenarios. The estimation results are shown in Figure 2-14. We observe that after increasing or decreasing flight demand by 20%, the average arrival delays also increase or decrease accordingly, matching our intuition that higher demand means more serious congestion. Moreover, some airports suffer substantial increased delays when flight demand is increased, such as ATL, EWR and ORD. For the entire network, the average arrival delays for the airport network are 23.24, 36.32, 16.83 minutes for the three scenarios, respectively. This reflects the well-known non-linear relationship between demand and delay in queueing theory, namely that delays are concave as the demand increases. However, our airport system is far from a simple queueing system as it includes many complicating factors, such as the initial delays, delays from other airports beyond 34 major ones, and the propagation delays across airports.

Experiment 2: We consider three scenarios having different levels of airport capacities. The first one is still the actual scenario with the actual flight demand and the actual airport capacities. We construct the second scenario without changing flight demand at any airport, but increase the airport capacity at each airport by 20%. With the same method, we construct the third scenario under the same conditions as the second scenario except decreasing the capacity by 20%. We then use the RAND model to estimate the average arrival delays based on these three scenarios. The estimation results are shown in Figure 2-15, where the blue bar represents the average arrival delay under the actual capacities, while the green and the red bar represent the average arrival delays with capacities of more than 20% and less than 20% compared to the actual scenario, respectively. As expected, the average arrival delays change in the opposite direction of the change of airport capacities. Aggregately, the average

Figure 2-14: Average arrival delays on Feb 14 2007 based on different flights demand levels

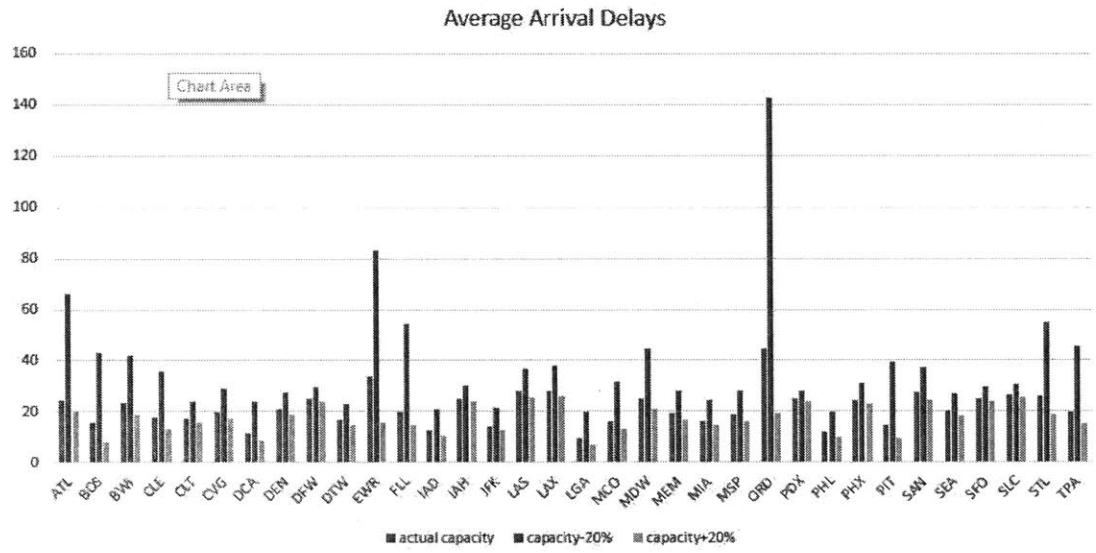


arrival delays for the whole airport network are 23, 18 and 43 minutes for the actual scenario, +20% scenario and −20% scenario, respectively. From Figure 2-15, we see that among all the airports, ORD is especially sensitive to the change of capacity as well as the three airports within the New York Metropolitan area (EWR, JFK and LGA). We conclude that flight delays are more sensitive to capacity changes than to demand changes since most of the airports have a more substantial flight delay increment compared to Experiment 1.

2.7 Cancellation Algorithms

The integrated model provides a framework in which different scenarios can be tested. We focus here on testing different cancellation strategies and evaluating their impacts on flight and passenger delays in the entire network. We consider five cancellation strategies for our computational experiments: the random cancellation strategy; the QE-based maximum-delay cancellation strategy; the QE-based maximum-ratio cancellation strategy; the RAND-based maximum-delay cancellation strategy; and the RAND-based maximum-ratio cancellation strategy. Each of these is described below. Generally, these algorithms cancel flights iteratively until they reach a pre-defined

Figure 2-15: Average arrival delays on Feb 14 2007 based on different airport capacities



number of cancelled flights. For each iteration, all the algorithms first cancel a flight based on a specific criterion, and then search for the consecutive flight of the cancelled flight so that these two flights corresponds to a round trip of the same aircraft. We denote this consecutive flight as the nearest flight of the current cancelled flight. It must be operated by the same airline and have the same tail number and aircraft type as the previous flight. In addition, this nearest flight has the same departure airport as the arrival airport of the previous flight or the same arrival airport as the departure airport of the previous flight. The algorithm for identifying the nearest flight is the same for all the cancellation strategies and is similar to the tail recovery algorithm: Flights are first sorted according to their airline, tail number and aircraft. We then find the flight whose scheduled departure time is later than but not too remote from the scheduled arrival time of the target flight. The acceptable time window is set to be between 20 minutes at a minimum and 24 hours at a maximum.

2.7.1 Random Cancellation Strategy

As shown in Algorithm 2, the random cancellation strategy randomly picks a flight from the flight pool, and cancels both that flight and its paired flight. The algorithm is very simple and fast, and turns out to give reasonably good approximate estimates of the passenger delays observed in reality (See 2.8).

Algorithm 2 The Random Cancellation Strategy

1: $N \leftarrow$ specified number of cancelled flights
2: $F \leftarrow$ flights from the integrated flight data set
3: N_1 flights that have been cancelled. $N_1 \leftarrow 0$
4: **while** $N_1 \leq N$ **do**
5: Randomly remove a flight f1 from F
6: Find the nearest flight f2, if one exists, with respect to f1 and remove it from F
7: $N_1 \leftarrow N_1 + (1 or 2)$
8: **end while**

2.7.2 QE-based Maximum-delay Cancellation Strategy

The QE-based Maximum-Delay cancellation strategy, as shown in Algorithm 3, first employs the Queueing Engine in each iteration of the algorithm, to compute the local delays at each airport. Only remaining flights (i.e. the flights that have not yet been cancelled) are considered when computing local delays. After summing up the local arrival delay and local departure delay for each flight estimated by the QE, the algorithm cancels the flight with the largest local delay, and its corresponding paired flight. In our computational tests, this strategy always performs poorly. It gives much larger passenger delay estimates, than in reality. The reason is that this strategy only considers local delays. An aircraft suffering a large delay in its first flight leg will not necessarily suffer a large delay in its second leg.

Algorithm 3 The QE-based Maximum-Delay Cancellation Strategy

1: $N \leftarrow$ specified number of cancelled flights
2: $F \leftarrow$ flights from the integrated flight data set
3: N_1 flights that have been cancelled. $N_1 \leftarrow 0$
4: **while** $N_1 \leq N$ **do**
5: Run QE for each airport and obtain local arrival delays and departure delays for all the flights in F
6: For each remaining flight, local delay$\leftarrow$ local arrival delay + local departure delay
7: Remove f1, the flight with the maximum local delay, from F
8: Find the nearest flight f2, if it exists, with respect to f1 and remove it from F
9: $N_1 \leftarrow N_1 + (1\text{or}2)$
10: Update demand of the QE for each airport
11: **end while**

2.7.3 QE-based Maximum-ratio Cancellation Strategy

The QE-based Maximum-Ratio Cancellation Strategy, shown in Algorithm 4 is almost the same as the QE-based cancellation strategy except that in Algorithm 4, we use the ratio between local delay and the number of seats on the aircraft as the cancellation criterion instead of just local delay. By dividing by the number of seats, the new criterion leverages the delays and seat capacity so that the algorithm tends to cancel flights with large delays but fewer passengers. In practice, Algorithm 4 outperforms Algorithm 3 in terms of alleviating passenger delays.

Algorithm 4 The QE-based Maximum-Ratio Cancellation Strategy

1: $N \leftarrow$ specified number of cancelled flights
2: $F \leftarrow$ flights from the integrated flight data set
3: N_1 flights that have been cancelled. $N_1 \leftarrow 0$
4: **while** $N_1 \leq N$ **do**
5: Run QE for each airport and obtain local arrival delays and departure delays for all the flight in F
6: For each remaining flight, local delay$\leftarrow$ local arrival delay + local departure delay
7: Divide the local delay of each flight by the number of seats of the aircraft performing that flight
8: Remove f1, the flight with the maximum ratio, from F
9: Find the nearest flight f2, if it exists, with respect to f1 and remove it from F
10: $N_1 \leftarrow N_1 + (1\text{or}2)$
11: Update demand of the QE for each airport
12: **end while**

2.7.4 RAND-based Maximum-delay Cancellation Strategy

The RAND-based Maximum-Delay Cancellation Strategy (Algorithm 5) is more sophisticated in terms of its within-loop algorithm. Instead of running only QE on each iteration, this algorithm executes the entire RAND model, considering flight delays for all the remaining flights within the entire airport network. The algorithm is time-consuming, but the payoff is that total flight delays are obtained for each aircraft itinerary rather than just local delays. Apart from the within-loop algorithm, all the other parts of the algorithm are the same as its QE version (Algorithm 3). The flights cancelled are those associated with the longest total itinerary delays.

Algorithm 5 The RAND-based Maximum-Delay Cancellation Strategy

1: $N \leftarrow$ specified number of cancelled flights
2: $F \leftarrow$ flights that from the integrated flight data set
3: N_1 flights have been cancelled. $N_1 \leftarrow 0$
4: **while** $N_1 \leq N$ **do**
5: Run RAND and obtain flight arrival delays and departure delays for all the flights in F
6: For each remaining flight, total delay of itinerary$\leftarrow$ total arrival delay of itinerary + total departure delay of itinerary
7: Remove f1, the flight with the maximum total itinerary delay, from F
8: Find the nearest flight f2 if exists with respect to f1 and remove it from F
9: $N_1 \leftarrow N_1 + (1\text{or}2)$
10: Update demand of the QE for each airport
11: **end while**

2.7.5 RAND-based Maximum-ratio Cancellation Strategy

Like its QE version, the RAND-based Maximum-Ratio Cancellation Strategy (Algorithm 6) computes the ratio between delays and seat capacity, but the delay used is the total itinerary delay rather than local queueing delays. The algorithm cancels the itinerary with the highest ratio. This algorithm turns out to be the best among the five in terms of alleviating passenger delays.

2.7.6 General comments

As explained above, all the cancellation strategies, with the exception of random cancellations, rely on estimates of flight delays by the RAND model, either partially by using QE or entirely by using both QE and PDA. Generally, for each iteration, these algorithms first find a flight to cancel in a greedy way based on a specific delay related index, such as the maximum delay or the maximum ratio between delay and seat numbers, and then search for another related flight to cancel in order to keep the

Algorithm 6 The RAND-based Maximum-Ratio Cancellation Strategy

1: $N \leftarrow$ specified number of cancelled flights
2: $F \leftarrow$ flights from the integrated flight data set
3: N_1 flights that have been cancelled. $N_1 \leftarrow 0$
4: **while** $N_1 \leq N$ **do**
5: Run RAND and obtain flight arrival delays and departure delays for all the flight in F
6: For each remaining flight, total delay of itinerary$\leftarrow$ total arrival delay of itinerary + total departure delay of itinerary
7: Divide the local delay of each flight by the number of seats of the aircraft performing that flight
8: Remove f1, the flight with the maximum ratio, from F
9: Find the nearest flight f2 if exists with respect to f1 and remove it from F
10: $N_1 \leftarrow N_1 + (1 \text{or} 2)$
11: Update demand of the QE for each airport
12: **end while**

entire airport network balanced. We can see that these algorithms attempt to develop logical strategies that may approximate airline thinking. In reality, for example, if a flight scheduled to fly from ORD to PHX is cancelled, its continuing flight leg from PHX to LAX should also be cancelled since this aircraft is not available. However, this does not mean all the flights operated by this aircraft should be cancelled. For instance, suppose an aircraft is scheduled to perform a round trip from ORD to ABE (Lehigh Valley Airport) and back in the morning and afternoon, followed by a round trip from ORD to ATW (Outagamie County Regional Airport) in the evening. If the ORD - ATW round trip is cancelled, this does not mean that the ORD - ABE round trip will also be cancelled. All the above algorithms can handle this situation since they only cancel round-trip flight legs in each iteration. If there is no round trip, the above algorithms will not cancel any flight leg.

However, a potential problem is that the algorithm cannot ensure that the flight itinerary will remain reasonable after each cancellation. For example[11], consider an aircraft that makes a few of round trips between ORD and PHX and then eventually flies from ORD to ABE, where it will stay at the end of the day. The algorithm may cancel only the second flight leg in this three-legged itinerary, which will then

[11] A real example from the data

decompose the itinerary to be from ORD to PHX and from ORD to ABE, leading to an unreasonable result (an infeasible itinerary). Luckily, this problem is not significant, as shown in Table 2.5. In this table, we compute the number of cancelled flights which lead to this unreasonable result under the Random Cancellation Strategy. The table shows that for each day we study, less than 5% of all flights are cancelled incorrectly[12] based on the RAND-based maximum-delay cancellation strategy. In other words, with the exception of these flights, the remaining flights have reasonable schedules. We have similar results for the other cancellation strategies. Therefore, though only cancelling at most two flight legs on each iteration, these algorithms prove to work reasonably well in generating feasible and logical flight schedules.

Table 2.5: Flights cancelled in the RAND-based maximum-delay cancellation strategy

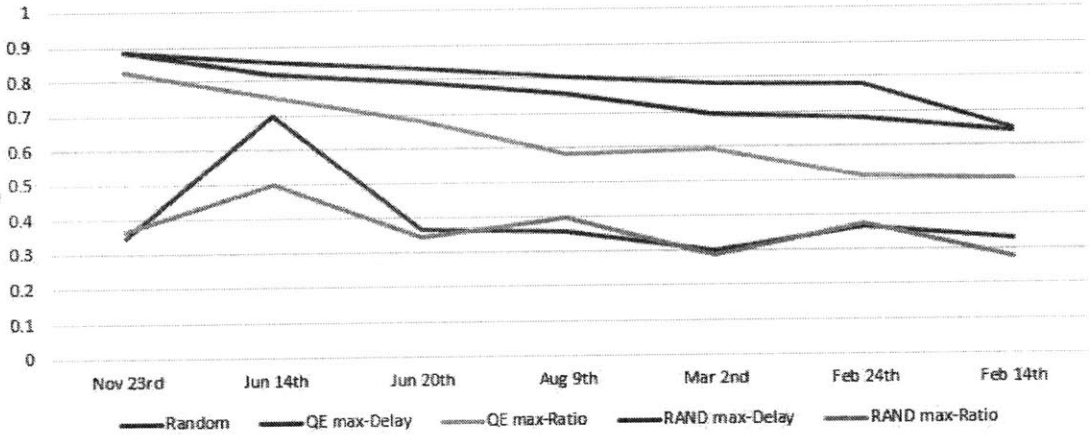
Date	Nov 23rd	Jun 14th	Jun 20th	Aug 9th	Mar 2th	Feb 24th	Feb 14th
Flights Cancelled	52	284	515	819	1238	2121	4385
Flights cancelled incorrectly (percentage)	1 2%	15 5.3%	11 2.1%	33 4.0%	37 3.0%	22 1.0%	53 1.2%

We test the five cancellation algorithms on the seven representative days. Figure 2-16 shows the percentage of the number of different aircraft cancelled out of the number of different flights cancelled for the five cancellation strategies over seven days. We observe that the more sophisticated cancellation algorithms tend to cancel more flight legs from a smaller group of aircraft. In other words, the number of aircraft involved in cancellations with sophisticated cancellation algorithms tends to be less. For example, the number of aircraft that RAND-based algorithms cancel is substantially smaller than those for QE based cancellation algorithms, while the number of aircraft that the random cancellations algorithm cancels is greater than that of RAND-based and QE-based algorithms. On the other hand, we also observe that for days with more delays, the same algorithm tends to cancel more flight legs from a smaller group of aircraft. For example, the QE-based max-ratio strategy

[12]Itineraries that separate an aircraft's trip into discontinuous flight legs

cancels flight legs from 0.5% of the aircraft on Feb 14th, while cancelling 0.8% of the aircraft on Nov 23rd. This observation matches the fact that, in reality, it is easier to cancel all of the flight legs associated with a specific aircraft than a subset of flight legs associated with several aircraft. In this sense, the RAND-based algorithms, and especially the RAND maximum-ratio based algorithm are quite reasonable.

Figure 2-16: Ratios of the number of aircrafts cancelled to the number of flights cancelled



Using the RAND model with these five cancellation algorithms, we estimate flight delays for five hypothetical scenarios. To compare flight delays for these five scenarios with those of the actual scenario, we plot five curves associated with the average arrival delays of these five hypothetical scenarios and one associated with the average arrival delay of the actual scenario over the seven representative days[13] in Figure 2-17. We summarize our observations as follows:

1. For days with small delays, such as Nov 23rd, Jun 14th and Jun 20th, the estimated flight delays in the hypothetical scenarios are similar to those in the actual scenario, whereas the difference becomes substantial for days with greater delays. For the days with substantial delays, most of the algorithms can reduce delays to different extents. For example, on Feb 14th, the RAND max-ratio strategy can reduce delay from 20 minutes to 13 minutes and on Aug 9th, the

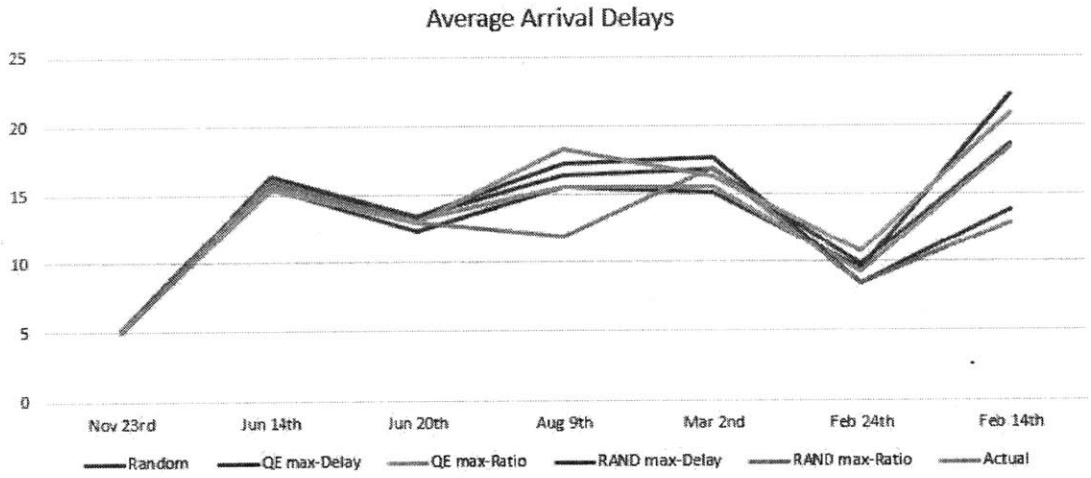
[13]See Table 2.1

QE max-delay strategy can reduce delay from 18 minutes to 15 minutes.

2. Among the five cancellation strategies, the RAND-based maximum-ratio cancellation strategy outperforms all the others, except for Mar 2nd. In particular, on Feb 14th and Aug 9th, it reduces delays by up to 30%.

3. In Figure 2-17, the curves generated by the two QE-based algorithms and the random algorithm have similar shapes, i.e., exhibit similar behavior.

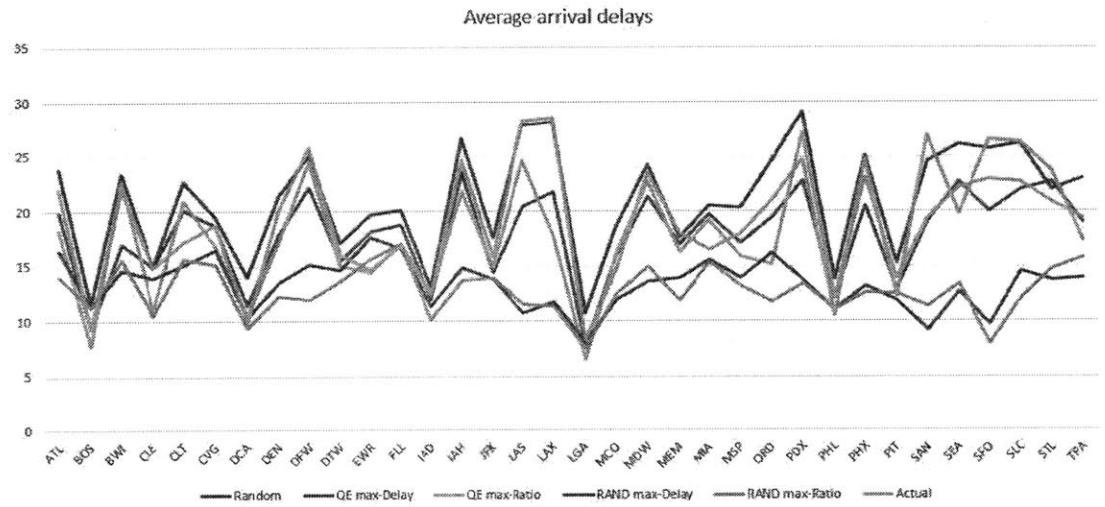
Figure 2-17: Average arrival delays for scenarios in which all five cancellation algorithms are applied



In order to study how flight delays are affected by cancellation strategies at the airport level, we present on Figure 2-18 average arrival delays on February 14th for each of the 34 major airports based on six scenarios (five based on cancellation strategies and one corresponding to the actual data). We observe that, instead of reducing delays, the QE-based maximum-delay cancellation strategy leads to more delays for most of the airports, whereas the QE-based maximum-ratio strategy can slightly reduce flight delays. The two RAND-based algorithms reduce the flight delays significantly.

Summarizing all of the figures and analysis presented above, we draw two conclusions in terms of flight delays:

Figure 2-18: Average arrival delays over 34 major airports on Feb 14th scenarios based on different cancellation strategies



1. The RAND-based strategies perform best and fit reasonably well the actual results because they consider itineraries over the entire network.

2. The performance of the QE-based strategies varies and depends on the criteria used.

2.8 Passenger Delays results and analysis

In this section, we present some results obtained through the full integrated model. The focus is on passenger delay metrics estimated by the model.

2.8.1 Passenger Delay metrics over six scenarios and seven days

We test six daily scenarios, including five hypothetical ones based on the five cancellation algorithms and the benchmark one based on Passenger Delay Calculator (PDC) [BFV10] and compute the estimated passenger delays for the seven representative days. The computational results with respect to different metrics of passenger delays

are shown from Figure 2-19 to Figure 2-23. The benchmark value is indicated as 'Actual' in these figures.

The passenger delay metrics computed in these figures are as follows:

1. Percentage of passengers with more than 15 minutes of delay.

2. Average delay per passenger: This metric is computed by dividing the total passenger delays by the total number of passengers.

3. Average delay per disrupted passenger: In the PDC, when a passenger misses a connecting flight or the connecting flight is cancelled, a passenger is labelled as disrupted. Average disrupted passenger delay is computed by dividing the total delays of disrupted passengers by the total number of disrupted passengers.

4. Percentage of disrupted passengers: This is the percentage of disrupted passengers in the total number of passengers on a specific day.

5. Percentage of disrupted passengers receiving default delays: The PDC will assign new flights to disrupted passengers according their priority. However, it may be impossible to assign some disrupted passengers to any flight. These are labelled as unassigned passengers when their waiting time reaches a pre-defined time threshold. Dividing the number of unassigned passengers by the number of total disrupted passengers, we obtain the percentage of disrupted passengers receiving default delays.

In Figure 2-19, we compute the percentage of passengers with more than 15 minutes delays over 7 days and 6 scenarios. From the figure, we observe that for days with more flights cancelled (from left to right), the percentage of passengers with more than 15 minutes of delay increases. The RAND max-ratio strategy reduces delays to a greater extent than the other strategies, whereas the QE max-delay strategy performs poorly for heavily congested days, but better for light-traffic days. The random strategy gives results similar to the actual scenario. The QE max-ratio and RAND max-delay strategies perform better than the QE max-delay strategy but worse than the RAND max-ratio strategy.

In Figure 2-20, we compute average passenger delays over 7 days and 6 scenarios. We observe that for days with more flights cancelled, the average passenger delays increase substantially. Among all strategies, the RAND max-ratio is still basically the best algorithm, while for heavily congested days, the QE max-ratio strategy also performs as well as the RAND max-ratio. The QE max-delay strategy is still the poorest strategy, whereas the RAND max-delay, and the random strategies have similar performance as the actual case.

In Figure 2-21, we compute the average delays of disrupted passengers over 7 days and 6 scenarios. We observe that the traffic load for different days does not lead to significant differences in terms of average disrupted passenger delays. The random cancellation strategy is on average the best compared to the others, but it still leads to higher delay estimates than actual for light-traffic days (Nov 23rd, Jun 14th and Jun 20th). QE max-delay, QE max-ratio, RAND max-delay, RAND max-ratio lead to even higher delays than the actual case.

In Figure 2-22, we compute the percentage of disrupted passengers over 7 days and 6 scenarios. We observe that for days with more flights cancelled, the percentage of disrupted passengers increases substantially. As in Figure 2-19, QE max-ratio and RAND max-ratio are best among all the algorithms in terms of percentage of disrupted passengers, whereas the QE max-delay, the RAND max-delay and the random strategies perform worse than the actual case.

In Figure 2-23, we compute the percentage of disrupted passengers receiving default delays over 7 days and 6 scenarios. We observe that for days with more flights cancelled, the percentage increases substantially. QE max-ratio and RAND max-ratio are the best among all the algorithms since they produce the least percentage of disrupted passenger receiving default delays, except Aug 9th for the QE max-ratio strategy. For days with a heavy load, QE max-ratio outperforms RAND max-ratio. QE max-delay and RAND max-delay are the worst performers compared to the actual case in this figure.

Summarizing all the above, we draw several conclusions:

Figure 2-19: Percentage of passengers with more than 15 minute delays over 7 days and 6 scenarios

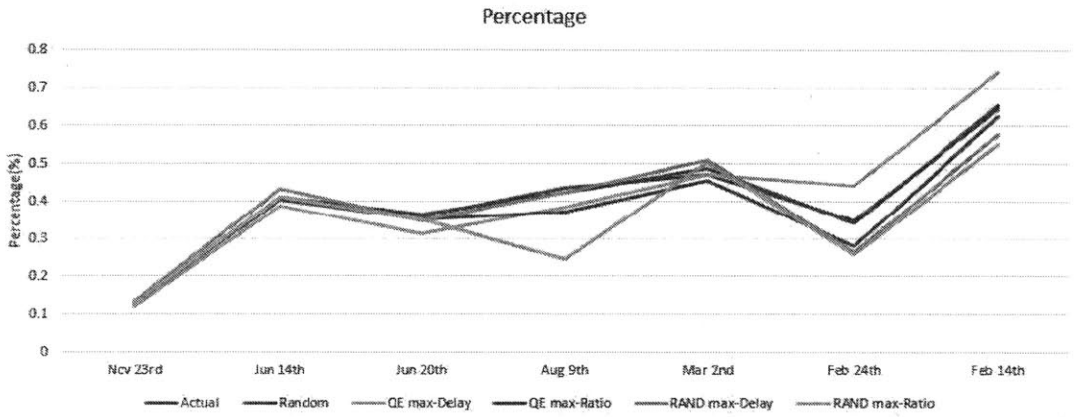


Figure 2-20: Average passenger delays over 7 days and 6 scenarios

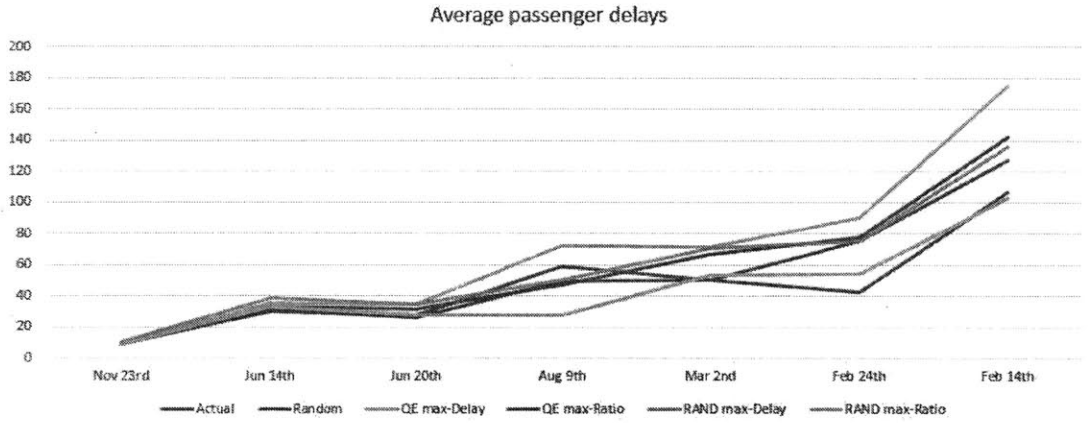


Figure 2-21: Average disrupted passenger delays over 7 days and 6 scenarios

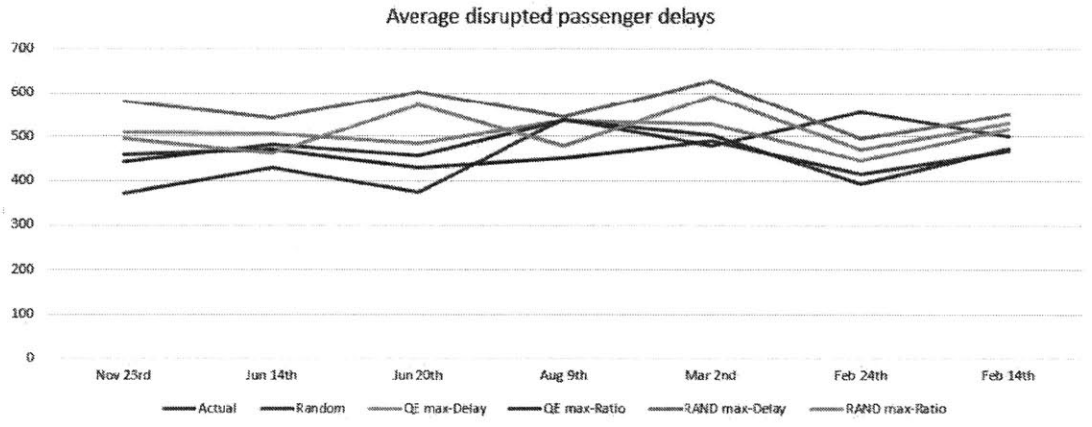


Figure 2-22: Percentage of disrupted passenger over 7 days and 6 scenarios

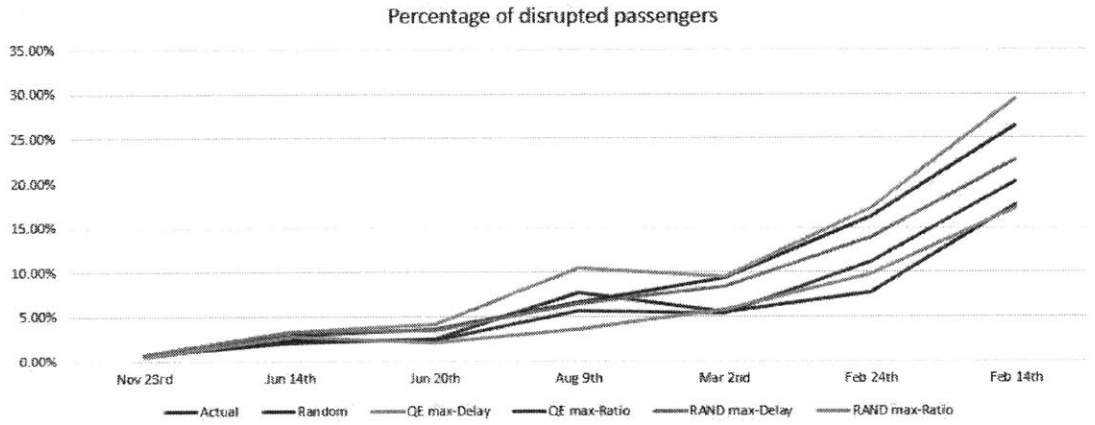
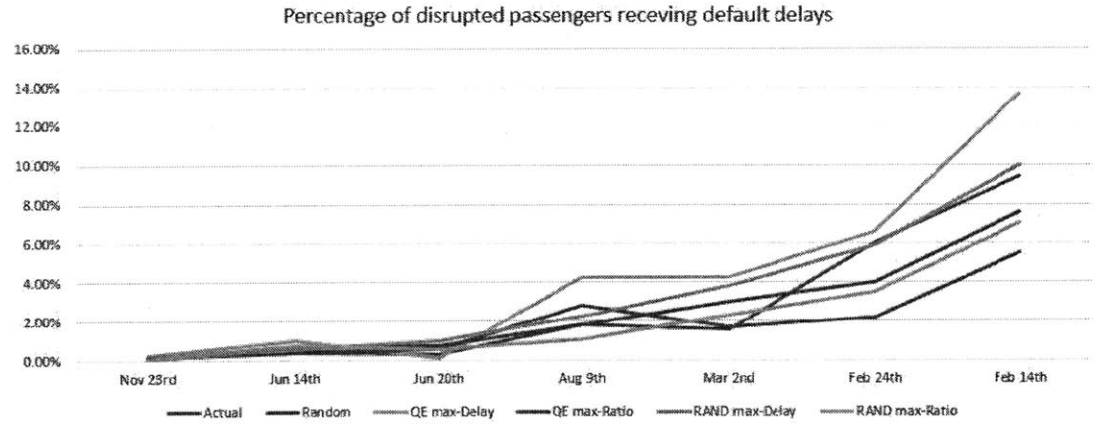


Figure 2-23: Percentage of disrupted passenger receiving default delays over 7 days and 6 scenarios



1. The RAND based max-ratio strategy outperforms others most of the time in terms of reducing passenger delays. This conclusion can be drawn from all the figures except the one for average disrupted passenger delays.

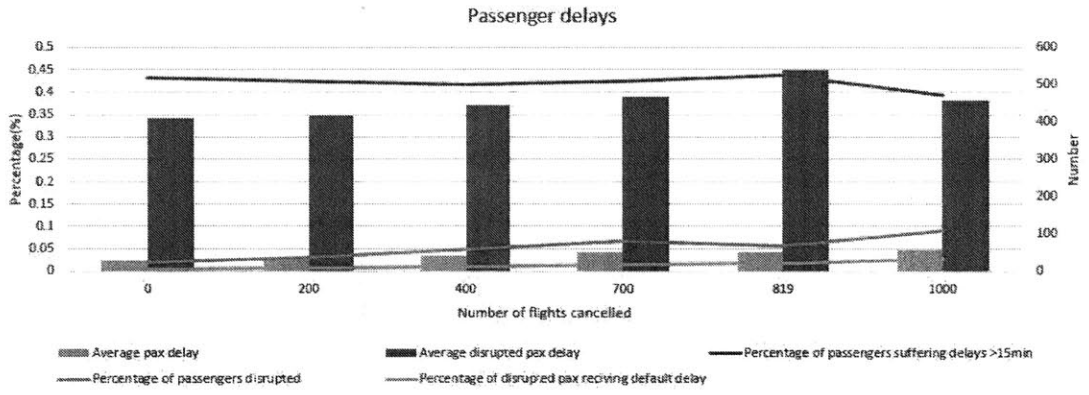
2. For days with fewer passenger delays (Nov 23rd, Jun 14th, Jun 20th), the cancellation algorithms have no significant impact on reducing passenger delays, whereas for those with heavy passenger delays, the two max-ratio strategies result in substantial improvement in terms of decreasing passenger delays.

3. The QE based max-delay cancellation strategy gives the worst results, when compared to the actual scenario. This once more shows that cancelling the flights with the maximum delay in each iteration will lead to bias.

2.8.2 Passenger delay metrics on Aug 9th given different number of cancelled flights

This experiment tests scenarios for August 9th, using the random cancellation strategy. The number of cancelled flights varies in the tests, so we can observe the impact of this number on the various metrics. In Figure 2-24, we consider six cases on Aug 9th, which are based on a different number of cancelled flights from 0 to 1000. In particular, the number of cancelled flights in these cases are 0, 200, 400, 700, 819 and 1000. We compute metrics including average passenger delays, average disrupted passenger delays, percentage of passengers suffering delays more than 15 minutes, percentage of passengers disrupted, and percentage of disrupted passengers receiving default delays for these six cases. In this Figure, we observe that with more flights cancelled, the average passenger delays, the percentage of passengers disrupted, and the percentage of passengers receiving default delays increases, which implies that flight cancellations do not necessarily reduce passenger delays, even though they may reduce some flight delays.

Figure 2-24: Passenger delays of scenarios with varying number of cancelled flights on Aug 9th



Chapter 3

The extended RAND model

In Chapter 2, we have discussed the integrated model that combines the RAND, the PDC and other models to estimate passenger delays under different scenarios. The integrated model focuses on a single day of operations and uses as its inputs that day's flight demand information, a real-time capacity profile for each airport, and the actual initial delays for flights scheduled to depart in the early time of a day. All these requirements prevent the model's application for predicting flight delays and passenger delays over a specified extended period of time that we are interested in, such as one month or one year. However, aviation administrators and planners are often more interested in estimates of aggregate flight or passenger delays in a future month or year, so they can initiate policies and actions aimed at alleviating delays. Therefore, we try to extend the integrated model to accommodate aggregate input data, such as monthly flight demand data. We build models that generate profiles of airport capacities and initial delay distributions over a long period of time, such as a month, and thus extend the flight delay estimation part of the RAND model to a similarly long period. We first briefly review in this chapter several widely used machine learning methods, including the logistic regression model, the Classification and Regression Tree (CART) model, and clustering methods. Then we proceed to construct capacity profiles either at a daily level or at an hourly level based on these models for all the 34 major airports by using airport capacities and weather condition data from the Airport Analysis Module of the Aviation System Performance Metrics

(ASPM). Section 3.3 presents results from the application of the aggregate RAND model that computes average monthly arrival delays at the airport level. We compare these estimates to actual delays from the ASPM database.

3.1 Model Review

3.1.1 Logistic Regression Model

The Logistic Regression Model is an important machine learning model which is widely used in classification problems. Consider a two-class classification problem, where we use $Y = 0, 1$ to represent the class label, and $\mathbf{X}$ to represent the feature vector or explanatory variables. Instead of directly predicting the value of Y, we try to find a regression model that predicts the probability of being in class 1 given explanatory variables, namely, $Pr(Y = 1|\mathbf{X} = \mathbf{x})$. This probability always takes a parametric form of the sigmoid function:

$$Pr(Y = 1|\mathbf{X} = \mathbf{x}) = \frac{1}{1 + e^{-\mathbf{W}^T\mathbf{x}}} \tag{3.1}$$

where $\mathbf{W}$ is the coefficient vector for explanatory variables.

In order to perform model selection for Logistic Regression, we compute McFadden's Adjusted R^2, which measures the relative quality of a statistical model for a given set of data and deals with the trade-off between the goodness of fit of the model and the complexity of the model. The formula is given by:

$$R^2_{adj} = \frac{ln\hat{L}(M_{Full}) - K}{ln\hat{L}(M_{Intercept})} \tag{3.2}$$

where $\hat{L}$ is the estimated likelihood given different models; M_{Full} represents the model we build with all the explanatory variables; $M_{Intercept}$ represents a naive model that only contains the constant term as an explanatory variable and K represents the number of parameters. By using the ratio between these two log-likelihood expressions with a penalty of K, the formula measures explanatory power by using the explanatory

variables in the model.

3.1.2 CART Model

The Classification and Regression Tree (CART) is a semi-parametric method which partitions the input space and uses different simple predictions in different regions of that space.

Consider again the classification case. The classifier is made up of

1. A partition function, π, mapping elements of the input space into exactly one of M regions, $R_1, \ldots, R_M$.

2. A collection of M output values, O_m, one for each region.

If we already knew a division of the space into regions, we would set $\hat{y}_m$, the constant output for region R_m to be the majority of the output values in that region; that is

$$O_m = \text{majority}_{\{i|x^{(i)} \in R_m\}} y^{(i)} \tag{3.3}$$

Define the error in a region as

$$E_m = |\{i|x^{(i)} \in R_m \text{ and } y^{(i)} \neq O_m\}| \tag{3.4}$$

Our objective is to select a partition to minimize

$$cM + \sum_{m=1}^{M} E_m \tag{3.5}$$

for some regularization constant c.

To achieve this goal, we establish a criterion (e.g. the misclassification error, the Gini index, etc [THF08]) based on the error defined in each region, for finding the best single partition of the data, and then apply it recursively to partition the space until the number of regions is greater than a threshold k. We will select the partition of the data that minimizes the sum of the errors of each partition.

Since the CART model tends to overfit the data, it might be tempting to regularize[1] by stopping after a somewhat large k, or by stopping when splitting a region does not significantly decrease the error.

Define cost complexity of a tree T, where m ranges over its leaves as

$$C_\alpha(T) = \sum_{m=1}^{|T|} E_m(T) + \alpha|T| \tag{3.6}$$

We will pick α using cross validation.

3.1.3 Cluster Methods

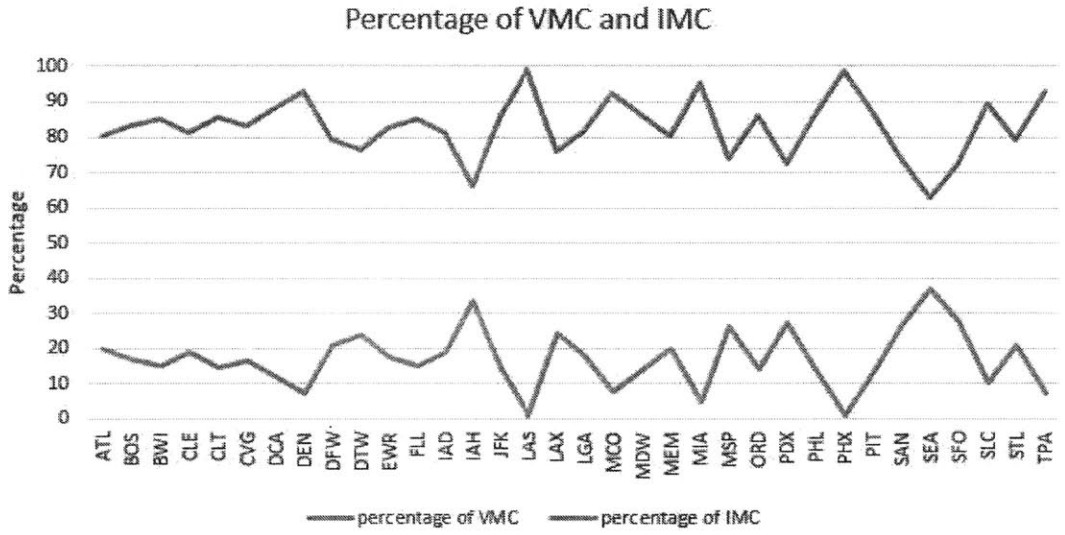
As a class of methods in supervised learning, Cluster methods do not need to specify the response variable and attributes. Instead, the cluster methods define some distance metrics for observations and select observations that are close to each other to construct clusters.

3.2 Construction of airport capacity profiles

As stated in section 2.1, the Airport Analysis Module from the ASPM is used for constructing Capacity Profiles for airports. The Airport Analysis Module contains weather conditions, including ceiling, visibility, temperature, wind angle, wind speed as well as expected hourly capacities for all the ASPM 77 airports. It also contains the Meteorological Conditions which we define as consisting of two states, either Visual Meteorological Conditions (VMC) or Instrument Meteorological Conditions (IMC). In the first part of this section, we will predict the hourly Meteorological Conditions based on weather conditions, by modelling the prediction problem as a classification problem. In particular, we will use the logistic regression model and the CART model. Then we will proceed to predict the airport hourly capacity directly, by using a linear regression model and a hierarchical cluster method. For all the models, we use data

[1]Regularization is a technique that prevents the model from overfitting.

Figure 3-1: Percentage of VMC and IMC hours at the 34 airports in 2007



of 2007 to estimate parameters in the model, and use data of January 2008 to test the accuracy of our predictions.

3.2.1 Logistic Regression for estimating meteorological conditions

We first show the percentage of hours under VMC and IMC for each of the 34 major airports in Figure 3-1. We observe that all the airports have a significant percentage of VMC and IMC hours except McCarran International Airport (LAS) and Phoenix Sky Harbor International Airport(PHX), which are in VMC practically all the time. Therefore, for most of the airports, the logistic regression model has an adequate number of observations from both levels of its response variable.

The best specification of the logistic regression model we find can be written as[2]:

$$V = \beta_{con} + \beta_{cei} CEILING + \beta_{vis} VISIBILITY + \beta_{temp} TEMPERATURE +$$

[2]CEILING is measured in Hundreds of Feet; VISIBILITY is expressed in statute miles, ranging from 0 to 99; TEMPERATURE is expressed in degrees Fahrenheit; WIND SPEED is in Knots; All the following tables in this chapter use these units for variables which represent weather conditions.

$$\beta_{wind} WINDSPEED + \gamma_1 TimeDummy_1 + \gamma_2 TimeDummy_2 + \gamma_3 TimeDummy_3$$

The probability of being in VMC or IMC is given as follows:

$$Prob(VMC) = \frac{1}{1 + e^{-V}}$$

$$Prob(IMC) = 1 - Prob(VMC)$$

where V is a scalar.

In this specification, β_{con}, β_{cei}, β_{vis}, β_{temp} and β_{wind} are coefficients for the corresponding weather conditions, whereas $\gamma_1, \gamma_2, and \gamma_3$ are coefficients for time-based dummy variables. In particular, we use the time interval from 6 pm to midnight as the benchmark and use dummy variables to represent other time intervals. $TimeDummy_1$ is 1 if the time of day of the observation is between midnight and 6 am, 0 otherwise; $TimeDummy_2$ is 1 if the time of day of the observation is between 6 am and noon, 0 otherwise, and $TimeDummy_3$ is 1 if the time of day of the observation is between noon and 6 pm, 0 otherwise.

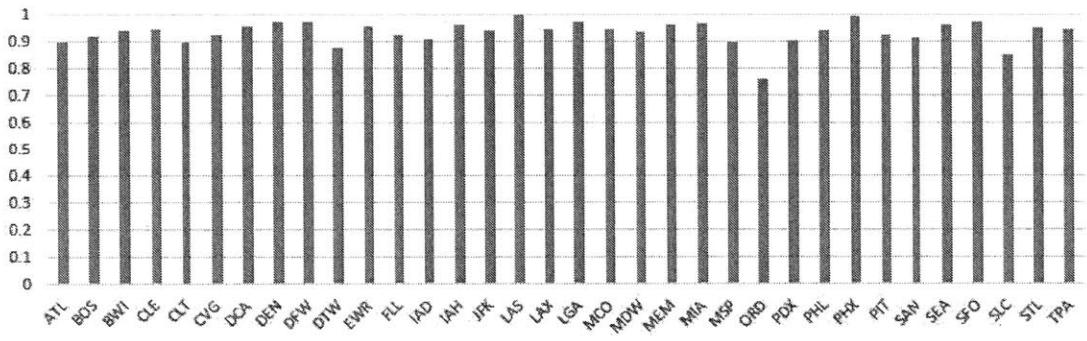
The parameter estimates and goodness of fit using data of 2007 are shown in Table 3.1. We observe that the goodness of fit in terms of McFadden's adjusted R^2 is quite good. When looking into the statistical test results of individual logistic regressions for each airport (not listed here), we find that the Z values for all the parameters are significant, except for LAS and PHX. As stated earlier, these two airports seldom have IMC observations, leading to the insignificance of the estimated coefficients.

From Table 3.1, we also observe that the coefficients for ceiling and visibility have positive signs, implying the higher the ceiling and visibility values, the greater the chance that an airport has a VMC hour. However, other factors such as temperature and wind speed contribute differently to predicting each airport's meteorological conditions. For instance, for ATL, a higher temperature generally means better weather conditions, thus leading to a VMC prediction for the airport, whereas the opposite is true for BOS. Furthermore, time dummy variables have different signs over different times of the day and different airports, suggesting that congestion at different airports

Table 3.1: Parameter estimation and Goodness fit

AIR-PORT	INTE-RCEPT	BETA CEI-LING	BETA VISIB-ILITY	BETA TEMP	BETA WIND-SPEED	GAMMA DUMMY1	GAMMA DUMMY2	GAMMA DUMMY3	McFa adj. R^2
ATL	-15.24	0.02	1.53	0.02	-0.13	-0.52	-0.35	0.41	0.70
BOS	-6.23	0.03	0.68	0.00	-0.02	0.08	0.42	0.43	0.69
BWI	-6.21	0.02	0.82	0.00	-0.02	-0.72	-0.29	0.25	0.69
CLE	-4.70	0.09	0.32	0.00	0.00	0.18	-0.11	0.11	0.63
CLT	-9.20	0.03	0.88	0.03	-0.08	-1.19	-1.42	-0.71	0.70
CVG	-5.69	0.04	0.45	0.03	-0.03	-0.23	-0.61	-0.30	0.65
DCA	-5.81	0.06	0.54	-0.01	0.00	-0.40	-0.08	0.44	0.73
DEN	-5.27	0.02	0.62	0.01	0.00	-0.51	-0.28	0.70	0.72
DFW	-11.97	0.05	0.99	0.02	-0.05	-0.99	-0.89	-0.08	0.73
DTW	-7.05	0.02	0.76	0.00	-0.01	-0.41	-0.39	-0.23	0.61
EWR	-5.21	0.05	0.52	-0.02	0.04	-0.39	-0.52	-0.02	0.69
FLL	-7.76	0.04	1.18	-0.06	-0.02	-0.40	-0.29	0.37	0.63
IAD	-10.34	0.01	1.27	0.00	0.03	-0.43	-0.10	0.23	0.71
IAH	-30.26	0.03	2.89	0.01	-0.05	-0.07	-0.38	0.44	0.71
JFK	-4.48	0.02	0.62	-0.02	0.02	0.00	0.29	0.25	0.62
LAS	-947.3	11.77	26.02	2.79	-2.67	-3.54	5.45	1.81	1.00
LAX	-4.95	0.04	1.06	-0.12	0.02	0.13	1.71	1.08	0.83
LGA	-7.07	0.06	0.64	-0.02	0.01	-0.14	-0.09	0.01	0.73
MCO	-4.70	0.04	0.42	0.01	-0.03	-0.51	-0.37	0.59	0.57
MDW	-4.17	0.05	0.44	0.00	-0.02	-0.24	-0.29	0.24	0.66
MEM	-15.23	0.05	1.17	0.02	-0.03	-0.53	-0.46	-0.01	0.77
MIA	-9.13	0.01	0.83	0.05	0.01	-0.29	-0.47	0.06	0.57
MSP	-21.98	0.03	2.18	0.00	-0.02	-0.47	-0.56	0.21	0.71
ORD	-4.61	0.03	0.48	0.01	0.02	-0.14	-0.54	0.02	0.63
PDX	-34.44	0.05	3.29	0.01	0.01	-0.38	-0.53	0.28	0.62
PHL	-4.07	0.03	0.54	-0.02	-0.01	-0.33	-0.27	0.30	0.64
PHX	-13.76	0.02	1.64	0.01	-0.02	0.29	0.70	0.36	0.76
PIT	-2.81	0.02	0.53	0.00	-0.01	-0.39	-0.56	0.02	0.57
SAN	-1.11	0.03	0.72	-0.12	0.11	0.17	1.07	0.37	0.66
SEA	-12.55	0.06	0.95	0.02	-0.06	-0.43	-0.33	0.29	0.63
SFO	-33.33	0.04	3.45	-0.06	-0.04	-0.46	0.48	0.80	0.82
SLC	-15.08	0.05	1.25	0.02	-0.01	-0.15	-0.39	-0.54	0.72
STL	-9.69	0.03	0.79	0.02	-0.04	-0.40	-0.62	-0.38	0.73
TPA	-3.50	0.02	0.58	0.00	-0.07	-1.32	-0.44	1.04	0.62

Figure 3-2: Prediction accuracy for the 34 airports of the logistic regression model



may happen in different patterns in terms of time of day. For several of the most congested airports in the U.S., such as ATL, LAX, JFK, DFW and DEN, the coefficients of the dummy variables are increasing with the time of the day, that is, $TimeDummy_1$ is less than $TimeDummy_2$, and $TimeDummy_2$ is less than $TimeDummy_3$, implying that the congestion level at these airports will tend to be more serious in the later part of a day.

To test the predictive power of the model, we apply these logistic regression models by using airport data from January 2008. Summing up correctly estimated observations (including observations of VMC that are estimated as VMC and observations of IMC that are predicted as IMC), we show a prediction accuracy for each airport in Figure 3-2. With almost all the airports we have more than 90% prediction accuracy. We conclude that the logistic regression model performs quite well in predicting VMC vs. IMC conditions.

3.2.2 The CART model for estimating meteorological condition

In this subsection, we use the CART model, in particular, the classification tree as an alternative way to predict meteorological conditions based on weather conditions. Since the CART model is a tree-based model, it is more intuitive than logistic regression in terms of mimicking the decision-making logic and process of human beings.

The best specification we have obtained contains two variables, ceiling and visibility, and it employs the Gini index as the node impurity measure.

Equipped with the CART model, we determine a corresponding classification tree (which is composed of values of some critical decision points, such as CEILING and VISIBILITY) for each of the 34 airports using 2007 data. Then we use January 2008 data to test the prediction accuracy of this approach. The result is summarized in Table 3.2. Each classification tree can be interpreted as a two-step decision process. Using ATL as an example, we first consider the ceiling parameter. If the ceiling is greater than 35.5 hundreds of feet, we claim ATL is in VMC, otherwise we base our decision on the visibility parameter. If visibility is greater than 6.5 miles, we believe ATL is in VMC, otherwise ATL is in IMC. This table implies that ceiling and visibility play more essential roles in terms of determining meteorological conditions than other weather conditions, such as temperature, wind speed and wind angle. This conclusion is consistent with our observations from the logistic regression model. Regarding the relative importance of ceiling versus visibility, we have no generic answer and conclude that their relative importance varies over airports. However, both models (classification tree first considers ceiling, then visibility or vice versa) perform quite well in terms of prediction accuracy using January 2008 data.

Additionally, if we compare the prediction accuracy of the CART model (the fourth column of Table 3.2) to that of the logistic regression (Figure 3-2), we find that they are very close differing only in the second digit. This implies that both models give very similar estimation results despite being based on different methodologies. This further confirms the applicability of machine learning techniques to this estimation task, which uses weather conditions to predict airport meteorological conditions.

3.2.3 Linear Regression Model for estimating hourly capacities

Since estimating airport meteorological conditions does not directly imply the level of capacity in an airport, we consider an alternative way to estimate capacity directly

Table 3.2: Critical decision points for classification trees and prediction accuracy

AIRPORT	CEILING (Hundreds of feet)	VISIBILITY (miles)	Prediction Accuracy
ATL	35.5	6.5	0.80
BOS	24.5	2.75	0.83
BWI	24.5	4.5	0.85
CLE	25.5	2.75	0.81
CLT	35.5	4.5	0.85
CVG	28.5	2.75	0.83
DCA	29.5	3.5	0.88
DEN	19.5	2.75	0.93
DFW	34.5	4.5	0.79
DTW	29.5	4.5	0.76
EWR	29.5	3.5	0.83
FLL	39.5	4.5	0.85
IAD	29.5	6.5	0.81
IAH	39.5	7.5	0.66
JFK	19.5	3.5	0.86
LAS	49.5	-1	0.99
LAX	24.5	2.75	0.76
LGA	31.5	3.5	0.82
MCO	24.5	2.75	0.92
MDW	18.5	2.75	0.86
MEM	49.5	4.5	0.80
MIA	19.5	4.5	0.95
MSP	34.5	7.5	0.74
ORD	18.5	2.75	0.86
PDX	34.5	7.5	0.73
PHL	22.5	3.5	0.86
PHX	32	6.5	0.99
PIT	17.5	2.75	0.87
SAN	19.5	2.75	0.73
SEA	39.5	2.75	0.63
SFO	34.5	7.5	0.72
SLC	52.5	2.5	0.90
STL	49.5	4.5	0.79
TPA	20.5	2.75	0.93

by using a linear regression model. The true capacity suggested by FAA is obtained from the Airport Analysis Module in the ASPM database. Our goal is to predict the capacity of each airport based on the weather information we have. Thus, the dependent variable is the airport capacity, whereas the independent variables include weather conditions, and dummy variables for time of day. The coefficient estimation results and goodness of fit are shown in Table 3.3. We observe that R^2 ranges roughly from 0 to 0.45, with 6 airports more than 0.3. Based on these results, we conclude that a linear regression model can explain some causality in terms of signs of coefficients. However, lacking a good enough goodness of fit, we do not have enough confidence to further use the model for computing flight delays. Comparing to the models built for estimating the meteorological conditions, capacity estimation requires factors beyond weather conditions.

3.2.4 Cluster Methods for estimating hourly capacities

In this subsection, we will employ a hierarchical clustering method to estimate airport capacity. We set the number of clusters to three. The reason for doing so is that we wish to compare our results to the capacity profiles suggested by the FAA. These profiles also identify three capacity levels for each airport.

We consider the hourly arrival rate and hourly departure rate for each airport as a two dimensional vector, and use Euclidean distance as the metric that describes the degree of similarity between them. Then we apply a hierarchical clustering method to divide capacities into three groups, compute corresponding mean values within groups and use them as the three levels of capacities for each airport. Moreover, the percentage of observations for each group is used as an estimator of the probability that a random observation will provide a capacity belonging to that specific group. The ultimate three-level capacity profile is shown in Figure 3-3 and Figure 3-4, where the blue, orange and gray curves represent the probability of high capacity, the probability of medium capacity, and the probability of low capacity, respectively. Comparing it to the three-level capacity profile suggested by the FAA shown in Figure 3-5, we observe that the range of capacity levels generated for each airport through this cluster method

Table 3.3: Estimation of parameters and Goodness of fit for Linear Regression Model with respect to capacity

AIR-PORT	INTE-RCEPT	BETA CEI-LING	BETA VISIB-ILITY	BETA TEMP	BETA WIND-SPEED	GAMMA DUMMY1	GAMMA DUMMY2	GAMMA DUMMY3	R^2
ATL	45.93	0.01	0.70	-0.02	-0.07	0.33	0.47	0.12	0.39
BOS	19.26	0.00	0.40	0.05	-0.18	-0.23	-0.73	0.31	0.17
BWI	19.73	0.00	0.12	-0.04	-0.01	-0.05	0.15	0.27	0.14
CLE	16.84	0.01	0.46	0.00	-0.09	0.10	0.18	0.21	0.21
CLT	27.71	0.00	0.44	-0.03	-0.08	-7.30	3.39	3.56	0.40
CVG	49.24	0.00	0.52	-0.07	-0.03	-3.41	2.59	2.75	0.18
DCA	15.84	0.00	0.10	0.00	-0.01	0.07	0.10	0.04	0.04
DEN	44.90	0.01	0.65	0.08	-0.17	0.16	0.91	-0.37	0.19
DFW	42.39	0.02	0.83	-0.03	-0.17	0.72	0.55	0.55	0.24
DTW	38.41	0.01	0.28	-0.07	-0.04	-0.36	-0.30	0.71	0.14
EWR	18.61	0.00	0.24	0.01	-0.08	0.55	0.49	0.08	0.23
FLL	17.05	0.00	0.13	0.02	0.02	0.12	-1.44	-0.21	0.26
IAD	28.74	0.00	0.28	0.01	-0.12	0.29	0.13	-0.16	0.08
IAH	35.98	0.01	0.41	0.00	-0.17	0.51	0.65	0.16	0.23
JFK	17.84	0.00	0.21	0.01	-0.05	0.11	0.19	0.27	0.17
LAS	22.30	0.00	0.30	0.02	-0.02	0.17	-0.05	-0.34	0.16
LAX	21.13	0.01	0.28	0.18	-0.05	0.73	-0.57	-0.70	0.25
LGA	20.60	0.00	0.06	-0.05	0.01	-6.00	0.41	0.77	0.29
MCO	40.45	0.00	0.02	-0.01	0.00	-0.02	-0.19	0.08	0.01
MDW	15.72	0.00	0.17	0.01	-0.06	0.03	0.16	0.14	0.08
MEM	32.77	0.01	0.30	-0.03	-0.06	-0.11	0.94	0.69	0.22
MIA	26.33	0.00	0.27	0.04	0.04	0.08	-1.10	-0.35	0.08
MSP	30.12	0.01	0.61	-0.03	-0.15	-0.38	0.92	1.11	0.19
ORD	37.77	0.01	0.97	0.01	-0.28	0.29	0.40	0.10	0.30
PDX	17.94	0.02	0.28	0.05	-0.04	0.65	-1.42	-0.71	0.35
PHL	20.12	0.01	0.44	0.00	-0.06	0.06	-0.18	0.35	0.27
PHX	26.56	0.01	0.33	0.02	-0.02	0.17	-0.34	-0.26	0.28
PIT	32.66	0.00	0.34	0.02	-0.02	0.41	0.12	-0.15	0.05
SAN	13.93	0.00	0.10	-0.04	0.00	0.01	0.07	0.09	0.12
SEA	19.62	0.00	0.19	0.01	-0.03	-0.09	-0.15	-0.03	0.21
SFO	12.73	0.01	0.27	0.07	-0.04	-2.22	2.58	2.91	0.47
SLC	24.43	0.01	0.75	0.01	0.01	0.22	-0.47	-0.23	0.22
STL	21.75	0.01	0.53	0.02	0.00	0.30	0.71	0.02	0.24
TPA	29.25	0.00	0.04	-0.03	0.01	-0.04	-0.29	-0.09	0.03

Figure 3-3: Three levels of 15-minute capacities identified by the Hierarchical Cluster Method

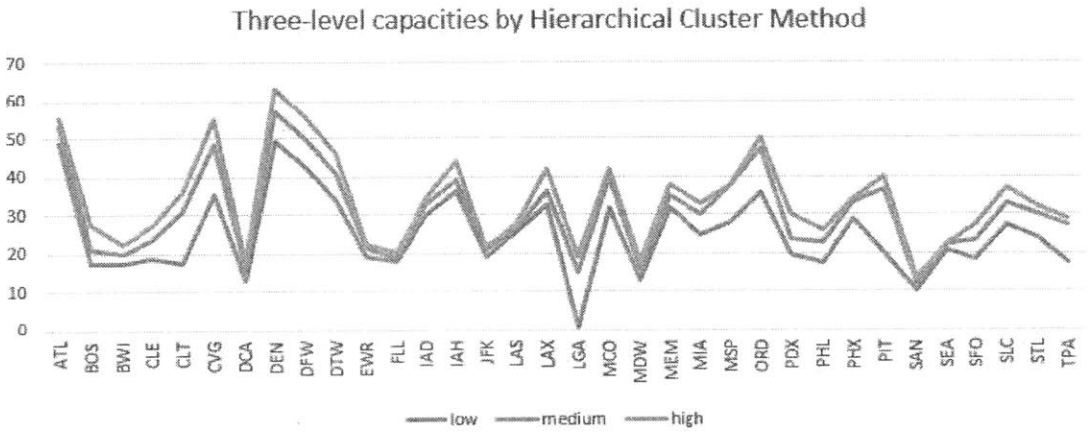
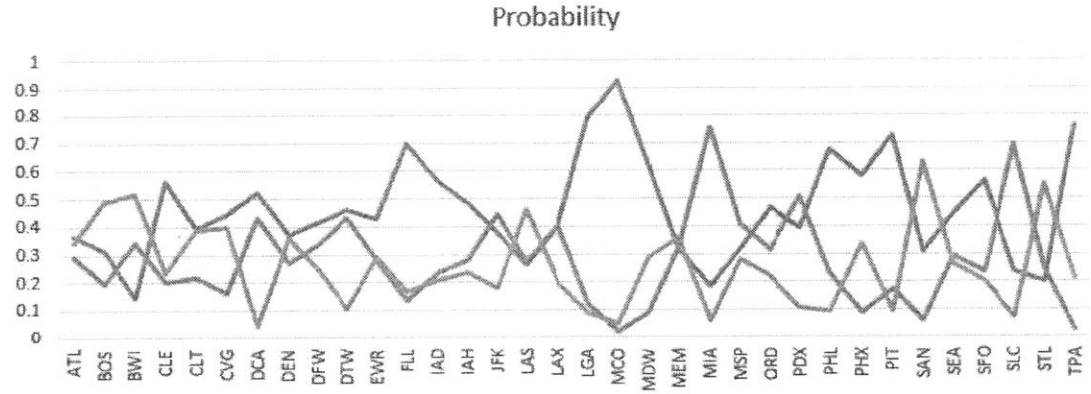


Figure 3-4: State Probability for the three levels of 15-minute capacities identified by the Hierarchical Cluster Method



is much wider than that of the FAA, and the state probability is more balanced. In contrast, the capacity profile given by the FAA (Figure 3-5) indicates that for more than roughly 80% of days the optimum capacity is available.

3.3 Estimation of Flight Delays and Discussion of Results

It was seen in Section 3.2 that Logistic Regression gives accurate estimates of VMC vs. IMC, while the CART model also provides similarly accurate estimates in a more intuitive way. However, when it comes to estimating capacities, the goodness of fit achieved by a linear regression model is only around 20-30%. Cluster methods are also not particularly successful in replicating the benchmark capacities provided by the FAA and their associated probabilities. Therefore, for estimating flight delays in January 2008, we only use the capacity profiles provided by the first two models. We denote these two capacity profiles as LogAvg and CARTAvg, respectively. Besides these two capacity profiles, we also consider two Benchmark capacity profiles produced by the FAA. These four capacity profiles are described in detail as follows:

1. LogAvg: The probability of the capacity values for this capacity profile is estimated through the logistic regression model for each hour based on weather conditions, while the capacity values are estimated using the mean of hourly capacities in either VMC group or IMC group from 2007 data.

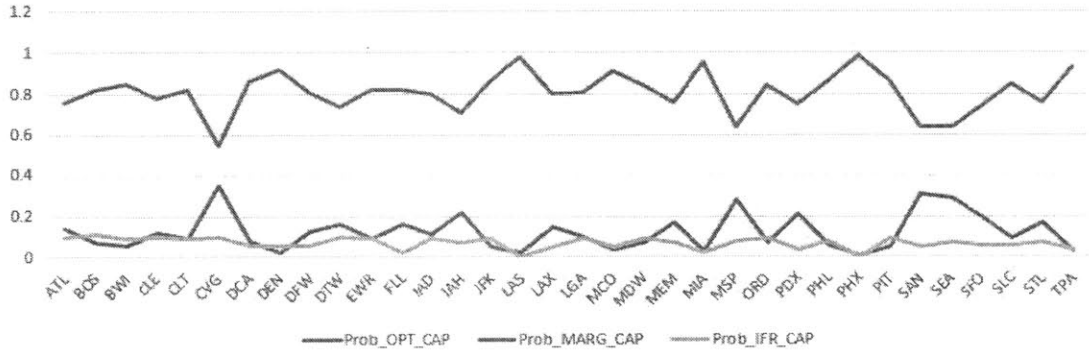
2. CARTAvg: The probability of the capacity values for this capacity profile is estimated through the CART model for each hour based on weather conditions, while the capacity values are estimated by the mean of hourly capacities in either VMC or IMC from 2007 data. The result is shown in Figure 3-6.

3. FAABen: This capacity profile is based on the Airport Capacity Benchmark Report in 2004[ACB04]. The detailed information involving the probability of the capacity levels and the magnitude of capacity levels is shown in Figure 3-5 and is also listed in Table A.1.

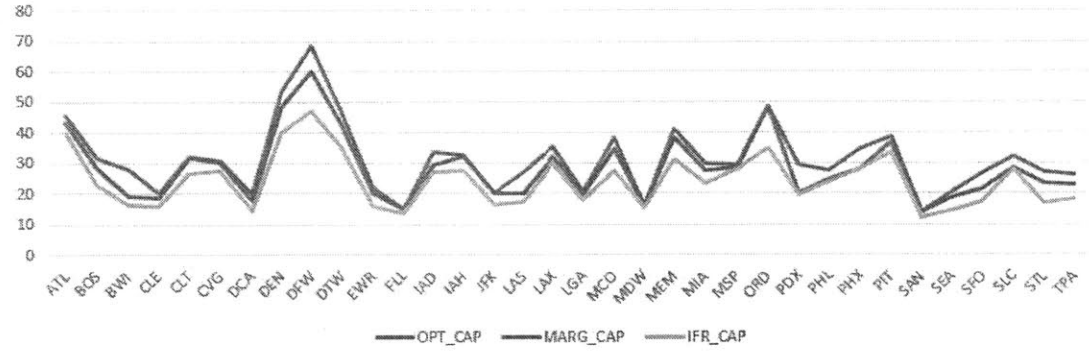
4. This capacity profile is derived from the Airport Capacity Benchmark Report in 2004. We convert the three levels of capacities to a single level of capacity by taking the expectation of capacities at all three levels from FAABen. The resulting capacity profile is as shown in Figure 3-7

Figure 3-5: FAABen Capacity Profile

(a) Probabilities for 15 minutes capacity values



(b) Capacity values



In addition to airport capacity profiles, another important input is the initial delays for early flight of each day. In Chapter 2, we directly use actual departure time for flights between 3 am and 6 am to replace their scheduled departure time, so that initial delays are taken into consideration. However, since we now consider an aggregate model based on monthly data, we relax this strong assumption by supposing an underlying distribution behind the initial delay of flights. As shown in Figure 2-12, initial delays for flights departing within 3 am to 6 am can be fitted by gamma distributions. The corresponding distribution parameter estimation results are listed in Tables 3.4 and 3.5 as well as test results from Kolmogorov-Smirnov goodness of fit tests.

Figure 3-6: Capacity values of Capacity profile CARTAvg(15 minutes capacities)

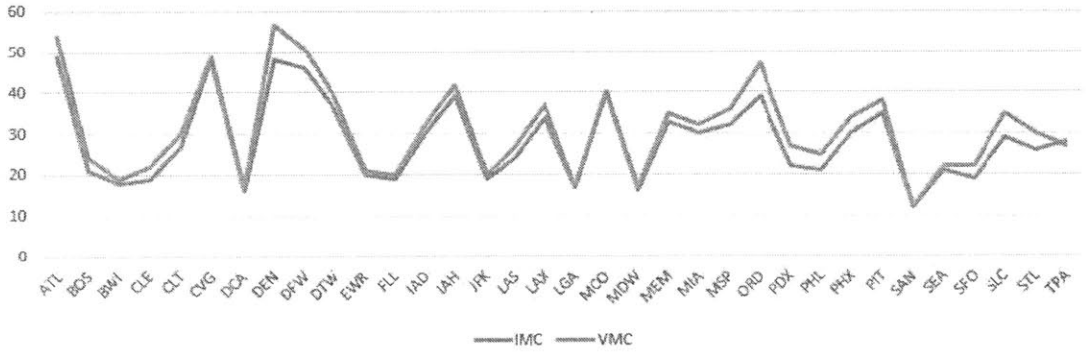


Figure 3-7: FAAAvg Capacity Profile (15 minutes capacities)

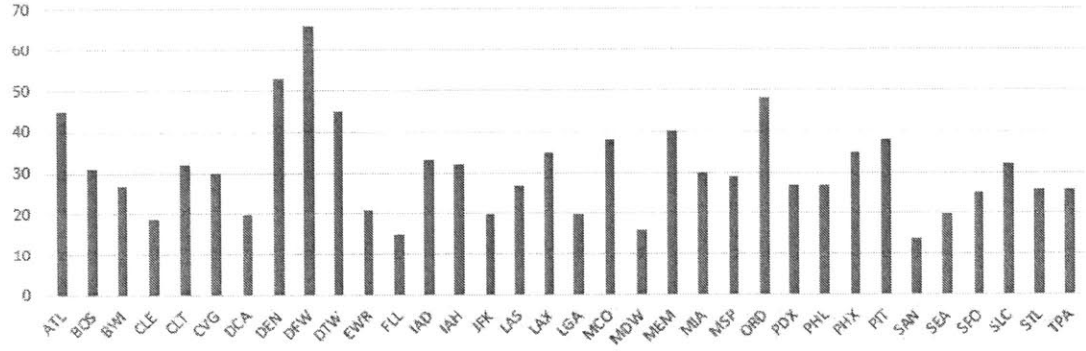


Table 3.4: Estimation and Goodness of fit test results for initial departure delays

Parameter Names	Parameter estimated	The estimated standard error
shape	0.721	0.0316
rate	0.0014	0.000075
D statistic	0.0011	
p value	0.581	

Table 3.5: Estimation and Goodness of fit test result for initial arrival delays

Parameter Names	Parameter estimated	The estimated standard error
shape	0.769	0.0350
rate	0.0013	0.000073
D statistics	0.0012	
p value	0.430	

Both of the results indicate that the estimates for the parameters are significant since the standard error is small. Moreover, the D statistic which represents the maximum distance between the CDF of data and the underlying true gamma distribution, is also small, and the corresponding p value is large, which implies that we cannot reject the hypothesis and conclude that the two distributions are significantly different. Therefore, we accept the hypothesis that these gamma distributions can be used as the underlying distributions for our data.

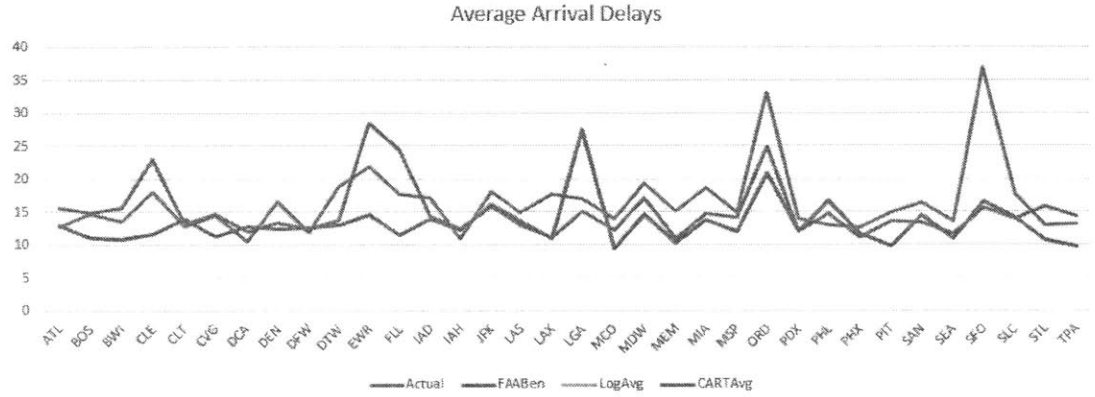
Based on the capacity profiles and underlying gamma distributions for initial delays, for each day of January 2008, we estimate hourly capacities directly using sampling capacity for each airport from the probabilistic daily capacity profiles. For the initial delays, we sample an initial delay for all the flights that are scheduled to take off between 3 am to 6am and use this initial delay to correct the scheduled departure time of the flight. We then use this corrected scheduled time information as the input for the extended RAND model. We compute average arrival delays for each airport for each day and further averaging them for 31 days. The final average arrival delays for the entire month for different capacity profiles are shown in Figure 3-8[3], as well as the benchmark result of actual average arrival delays computed directly from the ASPM database.

In this graph, the curve for LogAvg typically overlap with the curve of CARTAvg. In terms of numerical values, these two capacity profiles produce similar average capacities for all the airports. This observation is due to the fact that both models achieve close estimation accuracies. Readers can refer to Section 3.2.2.

We also observe that, for most airports, the delays generated by all four capacity profiles follow similar patterns even though they differ in magnitude. Moreover, all four capacity profiles can explain more than 70% of actual flight delays for most airports except EWR, FLL, ORD and SFO. Overall, this suggests that, since average flight delays on a monthly scale can be predicted at reasonable levels of accuracy, it is worthwhile to attempt to extend in the future the entire integrated model to a

[3]Note that LogAvg produce delays which are very close to that of CARTAvg, thus the curve for which is covered by that of CARTAvg and couldn't be seen from the graph

Figure 3-8: Average Arrival Delays for 34 airports in Jan 2008 using four capacity profiles



monthly scale and expect to obtain reasonable predictions of passenger delays.

Chapter 4

Conclusion

In this thesis, we have attempted to incorporate two models, the Airport Network Delay (AND) model and the Passenger Delay Calculator (PDC), within a single framework, so that the resulting new integrated model can compute passenger delays without requiring an actual flight-schedule input. The integrated model increases the usefulness and applicability of the PDC since it could be used with hypothetical scenarios, different flight cancellation strategies, etc.

We describe the framework of the integrated model for studying flight delays and passenger delays at a daily scale. The integrated model includes four components: a Tail Recovery Model, Flight Cancellation Algorithms, a Refined Airport Network Delay (RAND) model, and the PDC. The Tail Recovery Model recovers missing tail numbers for many flights recorded in the Aviation System Performance Metrics (ASPM) database. The Flight Cancellation Algorithms implement alternative strategies for flight cancellations in the presence of large delays, such as cancelling flights with long flight delays or flights with a large ratio of flight delay divided by the seating capacity of the aircraft. The RAND model is an extension of the AND, in which two implicit assumptions of the AND model have been modified. The RAND model produces better estimates of flight delays in the sense of replicating actual flight delays obtained from the ASPM database.

The overall integrated model is able to compute passenger delays and relies only on planned flight schedules rather than actual flight schedules. Moreover, the integrated

model facilitates the study of factors that influence flight delays, such as weather conditions and demand fluctuations, and evaluates the impact of different cancellation strategies on passenger delays. Using actual data from different days, we conclude that passenger delays can be reduced on the busiest traffic days through improved flight cancellation strategies.

In the second part of the thesis, we extend the RAND model to compute flight delays on a monthly scale using different capacity profiles as input. These capacity profiles can be directly obtained from Federal Aviation Administration (FAA) reports or constructed by using classical machine learning algorithms on airport-level data. We validate our estimation of flight delays by using data of January, 2008, showing that both the capacity profiles and the RAND perform well in terms of replicating the actual monthly flight delays. These results imply that an effort can be made to develop an integrated model incorporating the RAND, the PDC etc. at a monthly scale or even at any generic time scale.

Appendix A

Tables

Table A.1: Capacity per 15 minutes (and associated probabilities) of 34 major airports suggested by the FAA in the U.S.[ACB04]

IATA CODE	Name	OPT CAP (PROB)	MARG CAP (PROB)	IFR CAP (PROB)
ATL	Atlanta Intl	46(0.76)	43.25(0.14)	40(0.1)
BOS	Logan Intl	31.75(0.82)	28.75(0.07)	23(0.11)
BWI	Baltimore Washington Intl	28.25(0.85)	19.25(0.06)	16.5(0.09)
CLE	Cleveland Hopkins Intl	20(0.78)	18.75(0.12)	16(0.1)
CLT	Charlotte Douglas Intl	32.5(0.82)	32(0.09)	26.5(0.09)
CVG	Cincinnati/Northern Kentucky Intl	30.75(0.55)	30.5(0.35)	27.75(0.1)
DCA	Ronald Reagan Washington National	20(0.86)	18(0.08)	14.75(0.06)
DEN	Denver Intl	53.75(0.92)	48.5(0.02)	40.25(0.06)
DFW	Dallas/Fort Worth Intl	68.75(0.81)	60.25(0.13)	47.25(0.06)
DTW	Detroit Metropolitan Wayne County	46.75(0.74)	42.5(0.16)	35(0.1)
EWR	Newark Liberty Intl	22(0.82)	20(0.09)	15.75(0.09)
FLL	Fort Lauderdale Hollywood Intl	15.25(0.82)	15(0.16)	13.5(0.02)
IAD	Washington Dulles Intl	33.75(0.8)	29.25(0.11)	27.25(0.09)
IAH	George Bush Intl	32.75(0.71)	32.5(0.22)	27.5(0.07)
JFK	John F. Kennedy Intl	20.25(0.86)	20.25(0.05)	16.5(0.09)
LAS	McCarran Intl	27(0.98)	20(0.02)	17.5(0)
LAX	Los Angeles Intl	35.5(0.8)	32.25(0.15)	30(0.05)
LGA	LaGuardia	20.5(0.81)	19.75(0.1)	18(0.09)
MCO	Orlando Intl	38.5(0.91)	34.5(0.04)	27.5(0.05)
MDW	Midway Intl	16(0.84)	16(0.07)	15.5(0.09)
MEM	Memphis Intl	41.25(0.76)	38.5(0.17)	31.5(0.07)
MIA	Miami Intl	29.75(0.95)	27.75(0.03)	23.5(0.02)
MSP	Minneapolis Saint Paul Intl	29.25(0.64)	28.5(0.28)	28.25(0.08)
ORD	O'Hare Intl	48.75(0.84)	48.75(0.07)	35(0.09)
PDX	Portland Intl	29.5(0.75)	20(0.21)	19.5(0.04)
PHL	Philadelphia Intl	27.5(0.86)	24.75(0.06)	24(0.08)
PHX	Phoenix Sky Harbor Intl	34.75(0.99)	28.25(0.01)	28.25(0)
PIT	Pittsburgh Intl	39(0.86)	36.75(0.05)	33.75(0.09)
SAN	San Diego Intl	14.25(0.64)	14.25(0.31)	12.25(0.05)
SEA	SeattleTacoma Intl	20.5(0.64)	18.75(0.29)	14.5(0.07)
SFO	San Francisco Intl	26.75(0.74)	21.75(0.2)	17.5(0.06)
SLC	Salt Lake City Intl	32.5(0.85)	28.75(0.09)	28(0.06)
STL	LambertSt. Louis Intl	27.25(0.76)	23.25(0.17)	16.75(0.07)
TPA	Tampa Intl	26(0.93)	23(0.03)	18.5(0.04)

OPT CAP denotes Optimum Capacity; MARG CAP denotes Marginal Capacity; IFR CAP denotes Instrumental-Flight-Rules Capacity; PROB denotes probabilities associated to corresponding capacity class.

Table A.2: 43 other airports in the U.S.

IATA CODE	Name
ABQ	Albuquerque Intl
ANC	Ted Stevens Anchorage Intl
AUS	AustinBergstrom Intl
BDL	Bradley Intl
BHM	BirminghamShuttlesworth Intl
BNA	Nashville Intl
BUF	Buffalo Niagara Intl
BUR	Bob Hope
DAL	Dallas Love Field
DAY	Dayton International
GYY	Gary/Chicago Intl
HNL	Honolulu Intl
HOU	William P. Hobby
HPN	Westchester County
IND	Indianapolis Intl
ISP	MacArthur
JAX	Jacksonville Intl
LGB	Long Beach
MCI	Kansas City Intl
MHT	Manchester-Boston Regional
MKE	General Mitchell International
MSY	Louis Armstrong New Orleans Intl
OAK	Oakland Intl
OGG	Kahului
OMA	Eppley Air_x000C_eld
ONT	Ontario Intl
OXR	Oxnard
PBI	Palm Beach Intl
PSP	Palm Springs Intl
PVD	T. F. Green
RDU	RaleighDurham Intl
RFD	Chicago Rockford Intl
RSW	Southwest Florida Intl
SAT	San Antonio Intl
SDF	Louisville Intl
SJC	San Jose Intl
SJU	Luis Muoz Marn Intl
SMF	Sacramento Intl
SNA	John Wayne
SWF	Stewart Intl
TEB	Teterboro

Table A.3: 20 major airlines in the U.S.

IATA CODE	ICAO CODE	Name
AQ	AAH	Aloha Airlines
AA	AAL	American Airlines
AS	ASA	Alaska Airlines, Inc.
YV	ASH	Mesa Airlines
EV	ASQ	Atlantic Southeast(merged into ExpressJet after 2011)
XE	ASQ	ExpressJet
US	USA,AWE	US Airways
CO	COA	Continental Airlines
OH	COM	Comair(merged into Delta in 2012)
DL	DAL	Delta Air Lines
MQ	EGF	American Eagle Airlines
F9	FFT	Frontier Airlines
9E	FLG	Pinnacle Airlines
HA	HAL	Hawaiian Airlines
B6	JBU	JetBlue Airways
NW	NWA	Northwest Airlines(merged into Delta in 2008)
OO	SKW	SkyWest Airlines
WN	SWA	Southwest Airlines
FL	TRS	AirTran Airways
UA	UAL	United Airlines

Bibliography

[ACB04] Airport capacity benchmark report 2004. *U.S Department of Transportation, Federal Aviation Administration*, 2004.

[BFV10] C. Barnhart, D. Fearing, and V. Vaze. Modeling passenger travel and delays in the national air transportation system. 2010.

[Bra04] S. Bratu. An analysis of passenger delays using flight operations and passenger booking data. 2004.

[Cau93] Caulkins. The on-time machines: Some analysis of airline punctuality. *Operations Research*, 41(4):710–720, 1993.

[Hal99] W. Hall. *Efficient Capacity Allocation in a Collaborative Air Transportation System.* Phd thesis, Massachusetts Institute of Technology, Department of Aeronautics and Astronautics, 1999.

[Kiv74] P. Kivestu. *Alternative Methods of Investigating the Time Dependent M/G/k Queue.* Master's thesis, Massachusetts Institute of Technology, Department of French, 1974.

[Pyr12] N. Pyrgiotis. *A Stochastic and Dynamic Model of Delay Propagation Within an Airport Network For Policy Analysis.* Phd thesis, Massachusetts Institute of Technology, Department of Aeronautics and Astronautics, 2012.

[Shu95] R. Shumsky. *Dynamic statistical models for the prediction of aircraft take-off times.* Phd thesis, Massachusetts Institute of Technology, Department of Aeronautics and Astronautics, 1995.

[THF08] R. Tibshirani T. Hastie and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer, 2008.