METHOD FOR DERIVING MULTI-FACTOR MODELS FOR PREDICTING AIRPORT DELAYS

by

Ning Xu
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Operations Research

Committee:

_______________ Dr. Kathryn B. Laskey, Dissertation Co-Director

_______________ Dr. Lance Sherry, Dissertation Co-Director

_______________ Dr. Chun-Hung Chen, Committee Member

_______________ Dr. George L. Donohue, Committee Member

_______________ Dr. Clifton Sutton, Committee Member

_______________ Dr. Daniel Menascé, Associate Dean for Research and Graduate Studies

_______________ Dr. Lloyd J. Griffiths, Dean, the Volgenau School of Information Technology and Engineering

Date: 12/7/07 Fall Semester 2007
George Mason University
Fairfax, VA

Method for Deriving Multi-factor Models for Predicting Airport Delays

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

Ning Xu
Master of Science
George Mason University, 2004
Bachelor of Science
Dalian University, China 1993

Co-Director: Kathryn B. Laskey, Professor
Co-Director: Lance Sherry, Professor
Department of Systems Engineering and Operations Research

Fall Semester 2007
George Mason University
Fairfax, VA

DEDICATION

This is dedicated to my parents, my brother Tao, and sister Yi, but most of all, it is dedicated to God.

ACKNOWLEDGEMENTS

The creating of a dissertation can be an isolating experience, yet it is obviously not a lonely journey for me. My sincere gratitude goes to my advisors, professors, my family, and all my friends for their support, patience and love over the last few years.

First, I want to thank Dr. Kathryn Laskey, my mentor, who inspired and encouraged me to the study of the uncertainties. Her walking with me as I struggled and the guidance to recover when my steps faltered will never be forgotten. I am also thankful to her for carefully reading and commenting on countless revisions of this manuscript. She is one of the best advisors that a student can hope to find.

In that same vein, I want to thank Dr. Lance Sherry, my co-advisor. I have been amazingly fortunate to have an advisor who patiently supported and helped me overcome many crisis situations and finish this dissertation. I am deeply grateful for the insightful instruction he provided for my research. He taught me not only how to think from the strategic perspective and to write scientific papers, but also the value of perseverance.

I am indebted to Dr. Chun-hung Chen and Dr. George Donohue for their guidance and practical advice. They set high standards for the students and I am grateful to them for challenging me, and thus teaching me how to do research.

Dr. Clifton Sutton is one of the best professors that I have had in my life. His constructive comments and criticisms of my research were thought-provoking and they helped me make every step theoretically correct. He also kindly helped revise this dissertation to ensure correct grammar and consistent notation.

My parents and sister have all been encouraging and supportive to me, spiritually and financially. My brother even spent days to write some program trying to ease my load. The brothers and sisters in our fellowship, Zhao Tao, Elsa, Junhui, Qiyu, Zhuoting, Grace, and Kevin all assisted me in various ways during the course of my dissertation.

Finally, this acknowledgement would not be complete without the thanks to the researchers, especially Dr. Liya Wang, and students from the Center for Air Transportation Systems Research, and the students from the C^4I center, especially Andy Powell, for their comments on my presentation. Thank you Dr. Alexander Klein for providing the weather data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

OEP Operational Evolution Plan

FAA Federal Aviation Administration

ASPM Aviation System Performance Metrics

BTS Bureau of Transportation Statistics

NAS National Airspace System

AOC Airport Operation Center

ATS Air Transportation System

MIT Miles In Trial

IFR Instrument Flight Rules

VFR Visual Flight Rues

AAR Airport Arrival Rate (Airport Acceptance Rate)

ADR Airport Departure Rate

OIS Operational Information System

ITWS Integrated Terminal Weather System

TCWF Terminal Convective Weather Forecast

DOT Department of Transportation

BN Bayesian Networks

PPR Projection Pursuit Regression

MARS Multivariate Adaptive Regression Splines

CART Classification And Regression Trees

MSPE Mean Square Prediction Error

MTPE Mean Transformed Prediction Error

MATPE Mean Absolute Transformed Prediction Error

OLS Ordinary Least Squares

GARCH Generalized Autoregressive Conditional Heteroskedasticity

AIR21 Aviation Investment and reform Act for the 21st Century

LDA Linear Discriminant Analysis

Airport Codes:

ATL Atlanta Hartsfield International Airport, Atlanta, GA

BOS Boston Logan International Airport, Boston, MA

BWI Baltimore-Washington International Airport, Baltimore, MD

CLE Cleveland Hopkins International Airport, Cleveland, OH

CLT Charlotte Douglas International Airport, Charlotte, NC

CVG Cincinnati-Northern Kentucky International Airport, OH, KY

DCA Washington Reagan National Airport, Washington, DC

DEN Denver International Airport, Denver, CO

DFW Dallas-Ft Worth International Airport, Dallas-Ft Worth, TX

DTW Detroit Metropolitan Wayne County Airport, Detroit, MI

EWR Newark International Airport, Newark, NJ

FLL Ft Lauderdale-Hollywood International Airport, Ft Lauderdale, FL

HNL Honolulu International Airport, Honolulu, HI

IAD Washington Dulles International Airport, Washington, DC

IAH George Bush Intercontinental Airport, Houston, TX

JFK John F Kennedy International Airport, New York, NY

LAS Las Vegas McCarran International Airport, Las Vegas, NV

LAX Los Angeles International Airport, Los Angeles, CA

LGA La Guardia Airport, New York, NY

MCO Orlando International Airport, Orlando, FL

MDW Chicago Midway Airport, Chicago, IL

MEM Memphis International Airport, Memphis, TN

MIA Miami International Airport, Miami, FL

MSP	Minneapolis-St Paul International Airport, Minneapolis, MN
ORD	Chicago O'Hare International Airport, Chicago, IL
PDX	Portland International Airport, Portland, OR
PHL	Philadelphia International Airport, Philadelphia, PA
PHX	Phoenix Sky Harbor International Airport, Phoenix, AZ
PIT	Pittsburgh International Airport, Pittsburgh, PA
SAN	San Diego International Airport, San Diego, CA
SEA	Seattle-Tacoma International Airport, Seattle, WA
SFO	San Francisco International Airport, San Francisco, CA
SLC	Salt Lake City International Airport, Salt Lake City, UT
STL	Lambert-St Louis International Airport, St Louis , MO
TPA	Tampa International Airport, Tampa, FL

ABSTRACT

METHOD FOR DERIVING MULT-FACTOR MODELS FOR PREDICTING AIRPORT DELAYS

Ning Xu, PhD

George Mason University, 2007

Dissertation Co-Director: Dr. Kathryn B. Laskey

Dissertation Co-Director: Dr. Lance Sherry

Traffic Flow Management (TFM), in coordination with Airline Operation Centers (AOC), manage the arrival and departure flow of aircraft at the nations airports based on the airport Arrival and Departure rates for each 15 minute segment throughout the day. The management of traffic flow has become so efficient in the U.S., that approximately 95% of the delays now occur at the airports (not airborne). Inefficiencies in the traffic flow occur when non-traffic flow delays (e.g. carrier, turn-around, aircraft swapping and non-terminal area weather) are super-imposed on the traffic flow delays. Researchers have correlated these non-traffic flow delays at airports with sets of causal factors and have created models to predict aggregate delays at airports on the time scale of a day. To be consistent with the way traffic flow is managed, a model of causal factors of delays in

15 minute segments would provide the analytical basis for improving the efficiency of TFM.

This dissertation describes the development of multi-factor models for predicting airport delays in 15 minute segments at 34 OEP airports. The models are created using Multivariate Adaptive Regression Splines (MARS). The models, generated using historic individual airport data, exhibit an accuracy of 5.3 minutes for generated delay across all the airports, and 2.1 minutes for absorbed delay across all the airports. A summary of the factors that drive the performance of each airport is provided. The sensitivity of each of the factors is also analyzed.

Analysis of the models indicates that the factors that determine Airport Delays in 15 minute segments are unique to each airport. The most significant factors that generate delays at most of the nation's airports are Carrier Delay, GDP Delay at the outbound destination, and Departure Demand Ratio. Because of the relationship between these factors, and the propagation of delays throughout the network, the only way to mitigate system-wide delays is via a holistic network approach. The implications of these results are discussed.

The potential benefits from this research include providing: (1) researchers and analysts a method to identify systemic causes of delays in the NAS and study the trends of influential factors; and (2) airlines and Air Traffic managers a means to evaluate predicted delays while executing Traffic Flow Management initiatives.

CHAPTER 1

INTRODUCTION

1.1 Problem Background

The National Airspace System (NAS) is a large and complex stochastic system with thousands of interrelated components, such as administration, air traffic control centers, airports, airlines, aircraft and passengers (Donohue and Zellweger 2001). The complexity of the NAS creates numerous difficulties in air traffic management and control.

Among the most intractable of these problems is flight delay, with its high cost to airlines, complaints from passengers, and difficulties for airport operations. For the last ten years, the U.S. Air Transportation System (ATS) has experienced long delays and increasing costs. Among the 75 airports recorded in FAA ASPM database, the number of U.S. airports that experienced more than 20,000 hours of arrival delay per year increased from 27 in 1998 to 37 in 2005, and 39 in 2007 (in Figure **1.1**). The annual delay of 2007 is estimated by doubling the total delay from January to June 2007. Among these airports, Chicago O'Hare International Airport (ORD) and Hartsfield-Jackson Atlanta International Airport (ATL) had more than 100,000 hours of arrival delay in 2006.

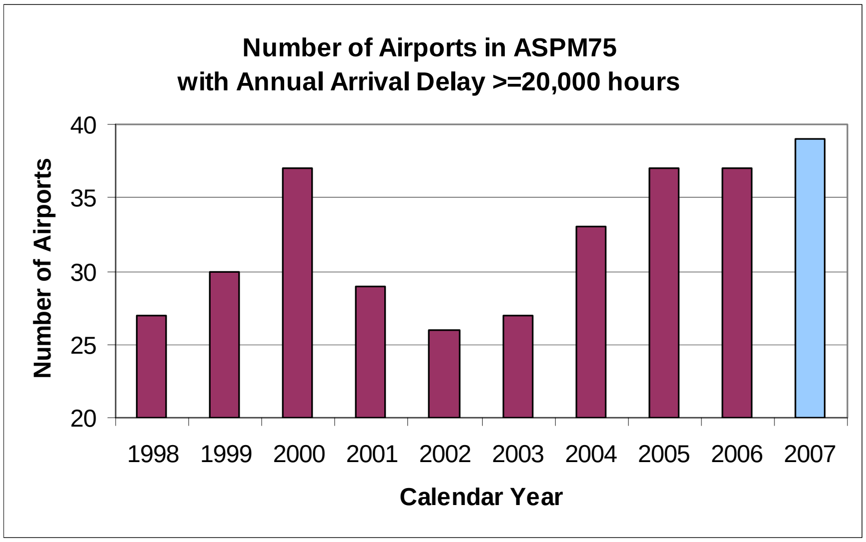


Figure **1.1**: Number of Airports Experiencing more than 20,000 Annual Arrival Delay. The value of delays of 2007 is estimated by doubling the total delay of the first 6 months of 2007.

Air traffic growth has put substantial pressure on the current ATS, and especially on capacity constrained areas of the air traffic infrastructure. According to Federal Aviation Administration (FAA) aerospace forecasts for fiscal years 2006-2017, the total combined instrument operations at airports with FAA and Contract Traffic Control Service are projected to increase from 48 million in 2003 to more than 64 million in 2017. Since, most major airports in the NAS are operating near capacity, ATS delays will increase as the demand reaches and exceeds the capacity limit.

1.1.1 Definition of Delays

The NAS can be modeled as a large scale network. The nodes of the network are the airports that serve to connect passengers to other flights or to transfer passengers and cargo to other modes of transportation. The arcs of the network are airways of the NAS.

A given aircraft originates and terminates flights at airports, traversing through several airports on its flight cycle throughput the day. At each node and on each arc of the network, an aircraft can accrue a delay or absorb a delay. Figure **1.2** illustrates generated delays and absorbed delays at airports and en route.

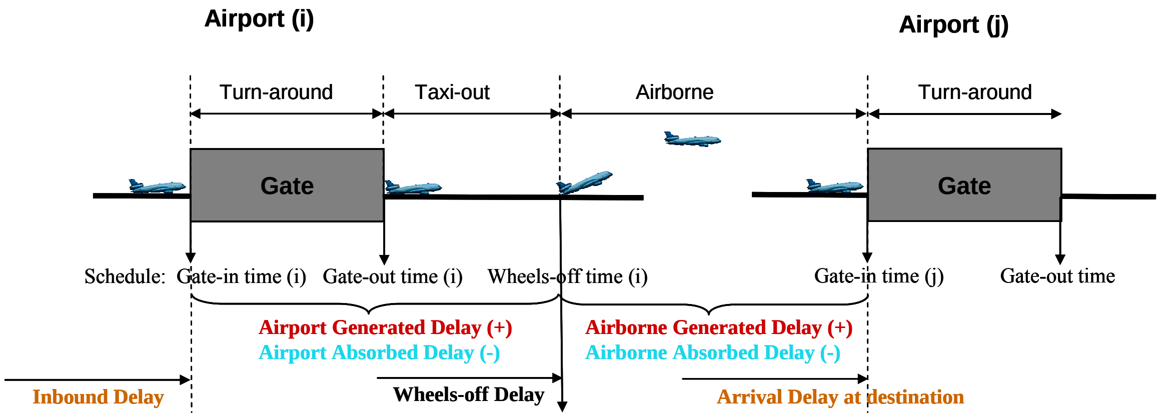


Figure **1.2**: Definition of Airport- and Airborne- Generated and Absorbed Delay

The positive delay that occurs at an airport is defined as Airport Generated Delay. This component of delay arises when an aircraft takes more time in any of its flight phases on the ground than is scheduled. Negative airport delay is defined as Airport Absorbed Delay which arises when an aircraft takes less time in any of its phases than is scheduled. The Airport Delay is the summation of Airport Generated Delay and Airport Absorbed Delay.

In the same manner, airborne delay is defined as Airborne Generated Delay if the delay is positive or Airborne Absorbed Delay if the delay is negative.

Besides airport delays and airborne delays, another critical delay in the NAS is inbound delay to an airport. Inbound delay is an accumulated value of delay on previous legs.

The relationships among the types of delays are summarized in Equation **1.1**.

$$
\begin{aligned}
\text{Airport Delay} &= \text{Airport Generated Delay} + \text{Airport Absorbed Delay} && (1)\\
\text{Airport Generated Delay} &= \max(0, \text{Turn-around Delay}) + \max(0, \text{Taxi-out Delay}) && (2)\\
\text{Airport Absorbed Delay} &= \min(0, \text{Turn-around Delay}) + \min(0, \text{Taxi-out Delay}) && (3)\\
\text{Wheels-off Delay} &= \text{Inbound Delay} + \text{Early\_arrival Gap} + \text{Airport Delay} && (4)\\
\text{Airborne Delay} &= \text{Airborne Generated Delay} + \text{Airborne Absorbed Delay} && (5)\\
\text{Arrival Delay at Dest.} &= \text{Inbound Delay} + \text{Early\_arrival Gap} \\
&\quad + \text{Airport Delay} + \text{Airborne Delay} && (6)
\end{aligned}
\qquad \mathbf{1.1}
$$

Two components of airport delay are turn-around delay and taxi-out delay. Taxi-out delay is the difference between the actual taxi-out time (from gate pushback to wheels-off) and unimpeded taxi-out time. In ASPM data dictionary, the unimpeded taxi-out time is calculated using a statistical function when the queues are of a minimal length based on aircraft queue lengths by carrier and airport (ASPM 2002).

Turn-around delay is the difference between actual turn-around time and scheduled turn-around time. However, the extra turn-around time for flights which arrived early from previous leg but departed on time or early on the next leg is not penalized as Generated Delay. For these flights, the turn-around delay is defined as zero and the discrepancy between the actual and scheduled turn-around time is defined as Early-arrival Gap.

The wheels-off delay at an airport is the summation of Inbound Delay, Airport Generated Delay, Airport Absorbed Delay and Early-arrival Gap (if the flight arrived early during a previous leg).

An example to calculate delays of a flight is provided in Figure **1.3**.

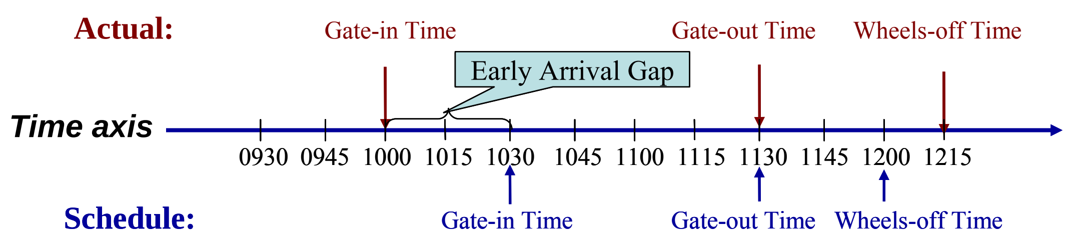


Figure **1.3**: An Example of Flight Delay Calculation

In this example, the flight arrived 30 minutes earlier than schedule, but departed on time. Hence, there was a 30-minute gap. The calculations of delays are as follows:

$$
\begin{aligned}
\text{Inbound Delay} &= \text{10:00-10:30} = \text{-30} \\
\text{Early Arrival Gap} &= 30 \\
\text{Turnaround Delay} &= 0 \\
\text{Taxiout Delay} &= 15 \\
\text{Airport Generated Delay} &= 0 + 15 = 15 \\
\text{Airport Delay} &= 15 + 0 = 15 \\
\text{Wheelsoff Delay} &= \text{Inbound Delay} + \text{Early Arrival Gap} + \text{Airport Delay} \\
&= \text{-30} + 30 + 15 \\
&= 15
\end{aligned}
$$

Taxi-in is one of the aircraft operations at an airport. For the purpose of this study, taxi-in delays are grouped with airborne delays for two reasons: (i) taxi-in delay is correlated with inter-arrival distances associated with the landing process. And (ii) taxi-in delay is a small, nearly constant portion of delays.

1.1.2 Characterization of Delays

Using Equation **1.1**, the delay statistics of Inbound Delay, Early Arrival Gap, Airport Generated Delay, Airport Absorbed Delay, Airport Delay, Airborne Generated Delay, Airborne Absorbed Delay, Airborne Delay and Arrival Delay at Outbound Destination (in Appendix Table **A.1**) were calculated for connecting flights in and out of 34 airports in the Operational Evolution Plan (OEP) using the data set of individual flights collected from the BTS Airline On-time Performance database (BTS June, July and August 2005).

These delays were summarized in Figure **1.4** in the order of the magnitude of Arrival Delay at Outbound Destination. The outbound destinations were restricted to 34 OEP airports. The total arrival delays at destinations from ATL, ORD and DFW are the top three among 34OEP airports. EWR, PHL, MSP and DTW are another group of airports having the second highest total arrival delays.

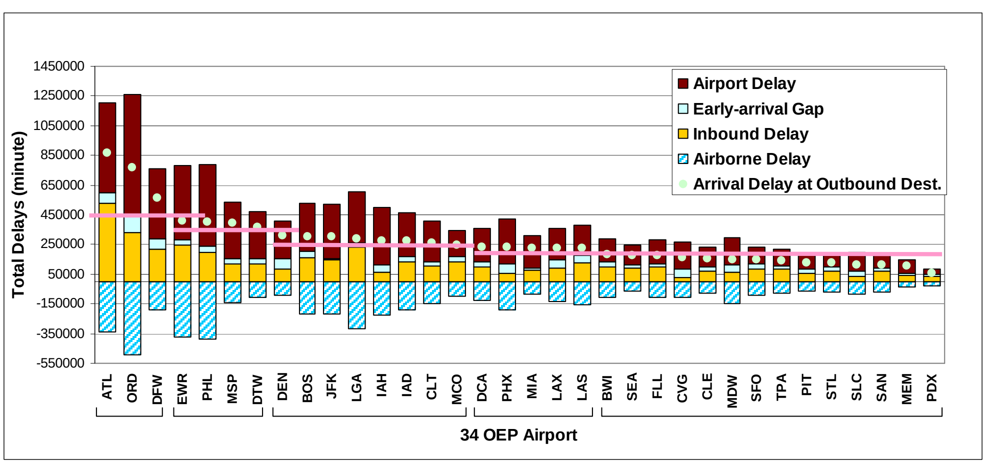


Figure **1.4**: Total Delays in Summer 2005 at 34 OEP Airports Ordered by Arrival Delay at Outbound Destination (minute).

Since the numbers of flights are different at each airport, the average delays per flight were compared (in Figure **1.5**). PHL, JFK, and EWR are the top three airports in the first group followed by ORD, MSP, MIA, IAD, IAH, and LGA. The average Airport Delay per flight at these airports is more than 15 minutes in summer 2005.

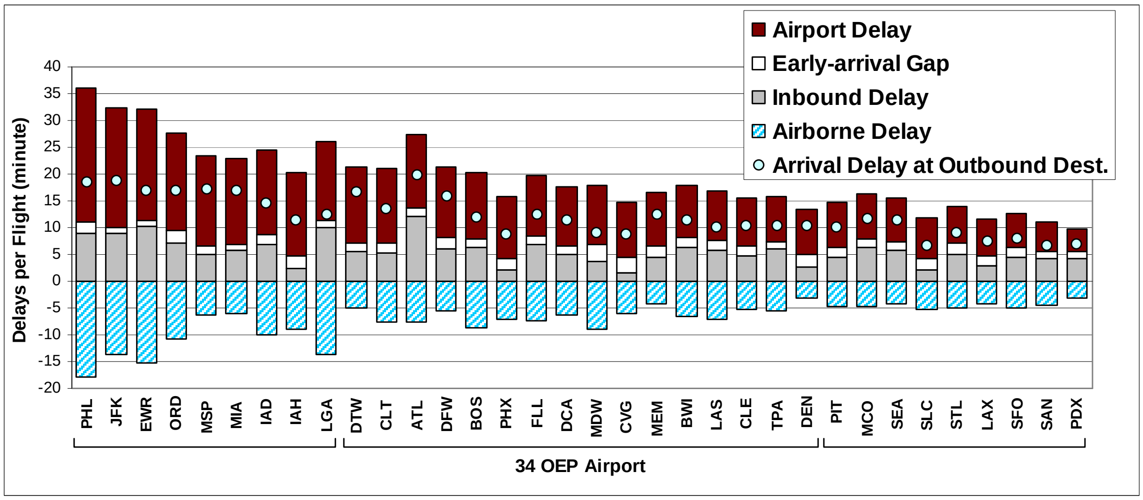


Figure **1.5**: Delays per Flight in Summer 2005 Ordered by Arrival Delay at Destination Ordered by Average Airport Delay per Flight (minute).

It can be seen from Figure **1.5** that PHL, EWR and LGA had very high airport delays but a great deal of these delays was absorbed in the airborne phase which can be attributed to the airline schedule padding. The breakdown of Arrival Delay at Outbound Destination is different for each airport.

Figure **1.6** shows the proportion of each type of delay to the total Arrival Delay at Outbound Destination at 34 OEP airports.

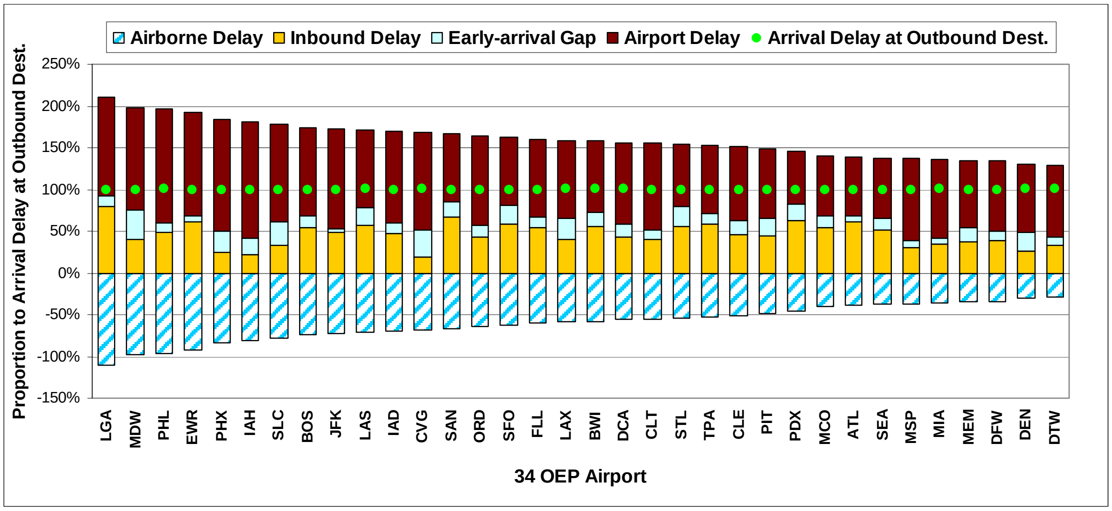


Figure **1.6**: Contributions of Airborne, Inbound, Early-arrival, and Airport Delay to Arrival Delay at Outbound Destination ordered by Airborne Delay (Percentage). The Inbound Delay accounts for 46%, Early Arrival Gap 14.5%, Airport Delay 98.6%, and Airborne Delay -59.1%.

From Equation **1.2** , it can be calculated that the total Inbound Delay of 34 OEP airports (69,648 hours) accounts for 46% of the total Arrival Delay at Outbound Destinations (151,253 hours), Early Arrival Gap accounts for 14.5%, and the Airport Delay (149,061 hours) accounts for 98.6%. The overall Airborne Delay (including taxi in delay at the destination) is a negative value (-89,427 hours). Its magnitude is 59.1% of the total Arrival Delay at Outbound Destinations.

$$\begin{aligned}\text{Arrival Delay at Outbound Destinations} = {} & \text{Inbound Delay} \\ & + \text{Early Arrival Gap} \\ & + \text{Airport Delay} \\ & + \text{Airborne Delay}\end{aligned} \qquad \mathbf{1.2}$$

When Mueller and Chatterji evaluated 21 days of data from October and November 2001 in Post Operations Evaluation Tool (POET) database, they found that 84% of all delays at ten U.S. airports occurred on the ground (2002). In the summer of year 2005, Airport Generated Delays accounts for 91.37% of the total Airport and Airborne Generated Delays at 34 OEP airports. This observation is consistent with the findings of Muller and Chatterji that the majority of flight delays are generated at airports. Since 1998, air traffic control initiatives such as Ground Delay Program (GDP) have been implemented to convert the unavoidable airborne delay to safer and cheaper ground delay at origin airport (Ball and Lulli 2004). Overall, the reliability and efficiency of the NAS is directly determined by the magnitude of these delays.

1.1.3 Characterization of Factors Related to Delays

Sussman states that the overall behavior of a complex transportation system is difficult to predict because the degree and nature of the relationships of the components of this system is imperfectly known (2004).The airport delays include delays resulting from carrier operations, congestion, weather, and delays by Ground Delay Programs and other traffic flow management initiatives. These delays are stochastic phenomena, and may be correlated to other factors of delays.

Existing research has also suggested that delays in the NAS result from many interrelated factors.

- Delays are caused by multiple causal factors (e.g. Allan et al 2001).

- Causal factors are correlated with other causal factors (e.g. Callaham et al. 2001). This creates problems when estimating the impact of individual causal factors.
- The inherently adaptive nature of the NAS creates strong non-linear relationships between causal factors (Donohue 2003). Delays propagate due to a highly connected schedule designed to maximize use of the expensive transportation resources (Beatty et al. 1998).

Mitigation of delays in the NAS requires an understanding of the factors that cause delays. An analysis method is required to accurately predict delays and to measure causal factors' contributions to a specific delay so that the appropriate corrective actions can be taken to prevent and alleviate delays.

1.2 Problem Statement

Traffic Flow Management (TFM), in coordination with Airline Operation Centers (AOC), manage the arrival and departure flow of aircraft at the nations airports based on the airport Arrival and Departure rates for each 15 minute segment throughout the day. The management of traffic flow has become so efficient in the U.S., that approximately 95% of the delays now occur at the airports (not airborne). Inefficiencies in the traffic flow occur when non-traffic flow delays (e.g. carrier, turn-around, aircraft swapping and non-terminal area weather) are super-imposed on the traffic flow delays. Researchers have correlated these non-traffic flow delays at airports with sets of causal factors and have created models to predict aggregate delays at airports on the time scale of a day. To be consistent with the way traffic flow is managed, a model of causal factors of delays in 15 minute segments would provide the analytical basis for improving the efficiency of TFM.

The research addresses the following problems:

- What are the direct causal factors that generate airport delay?
- What is the proper model to quantitatively estimate the impact of these factors on delays?
- Is the predictive model of airport delay valid?
- How should the degree of each factor's influence on airport delays be measured?
- What is the quantitative value of each factor's influences?

1.3 Research Scope

The airport delays at 34 airports in the FAA's Operational Evolution Plan (OEP) were investigated and modeled in this research. OEP 8.0 was established to reduce delay and meet future demand at the OEP airports by increasing the effective capacity of the NAS by 30% (FAA 1). 35 heavily operated airports were selected for the OEP. These airports account for 73% of total enplanements and 69% of total operations in the NAS (Bhadra and Texter 2005). Collectively, these airports are referred as "OEP-35". Figure **1.7** shows the OEP-35 airports on the U.S. map. The research focuses on all OEP-35 airports, except HNL.

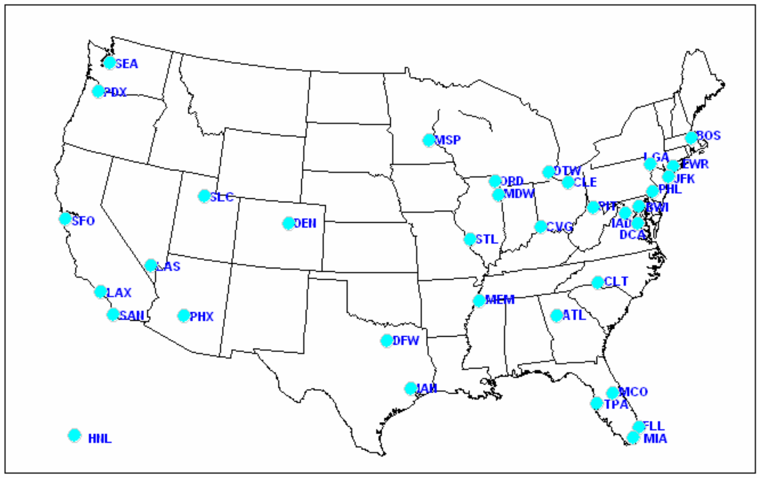


Figure **1.7**: OEP-35 Airport in the U.S.

The OEP-35 airports are highly connected. There are 1067 routes among the OEP-35 airports, of which 919 routes have at least one flight record per day during summer 2005 (from ASPM database). Figure **1.8** shows the number of destination airports each OEP-35 airport has. ATL, DEN, DFW, IAH, LAX and ORD have scheduled flights to all other OEP-35 airports. The majority of airports have more than 25 connections. HNL only has flights to nine OEP-35 airports. For this reason, HNL is not included in this research.

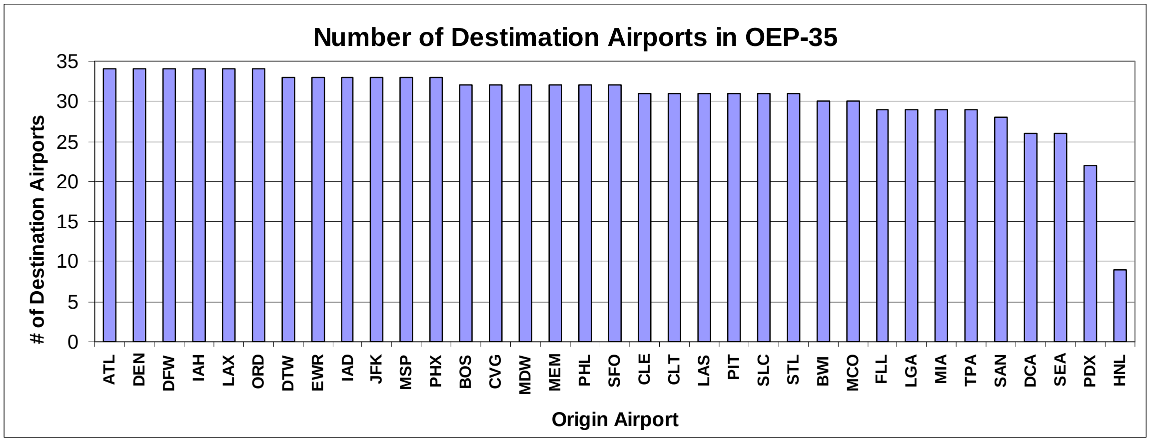


Figure **1.8**: Number of Routes Among OEP-35 airports

1.4 Objective

The objective of this research is to develop a method to predict the airport delay in the NAS. The method shall explicitly deal with complexity (multiple causal factors), collinearity (correlation between causal factors) and the nonlinear adaptive nature of the NAS. The method shall accept as inputs historical data on the daily performance of the NAS. The method shall generate as output:

(1) Multi-factor model for predicting Airport Generated Delay and Airport Absorbed Delay.

A multi-factor model represents the relationships between delays and the factors that cause them. Through the model, delays throughout a day can be predicted from the values of input factors. These predictions can be used to generate a graphic display like Figure **1.9**. In Figure **1.9** the output (at the top) reflects the magnitude of airport delay in each 15-minute epoch given the settings of input factors (listed at the bottom). A user can manipulate the input factors and see the resulting changes in predicted delay.

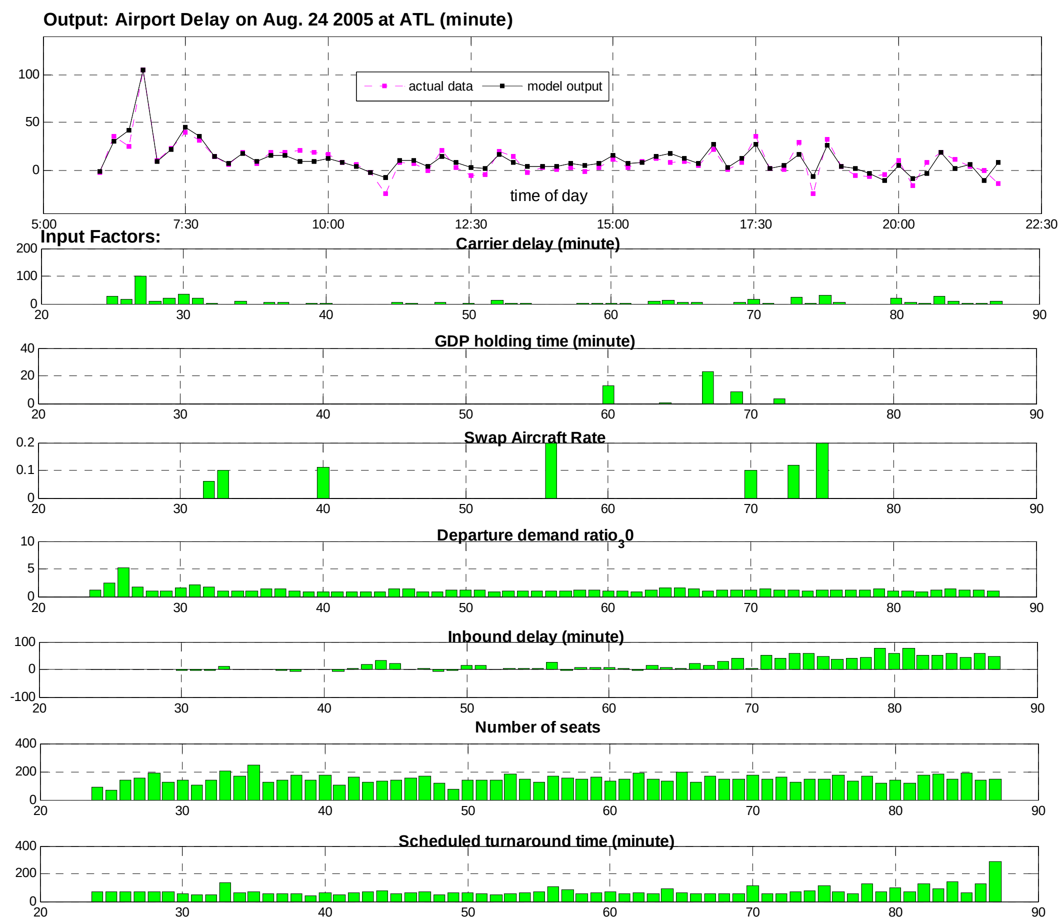


Figure **1.9**: Example of the Input and Output for a Multi-factor Airport Delay Model

(2) A multi-factor model trained with historical data, shall provide the user the ability to:

- Evaluate the degree to which causal factors contribute to a specific delay.
- Predict effects of interventions.

In this research, we assume the causal factors identified in the literature and by experts are the causes for delays. The objective of this research is to distinguish the factors which we can use to predict Airport Delay more accurately than with other factors.

1.5 Contributions

This research identifies (1) the important causes of delays and (2) the strength of each cause's influence on delay.

AOC and TFM personnel could apply the multi-factor models to predict airport delay with a tool, as illustrated in Figure **1.10**. Figure **1.9** shows the prediction of a single day given the values of inputs factors of each 15-minute epoch. Figure **1.10** plots the average Wheels off Delay and associated value of influence factors at LGA for each 15-minute epoch during the summer 2005. The setting of each factor can be adjusted by choosing the value from pull-down menus to the left of each factor's bar chart. Such a tool would enable personnel to perform "what if" analysis by making changes in causal factors at various times of the day and observing the predicted effects. The display would include multiple delay predictions to better understand the impact of one or more types of delays.

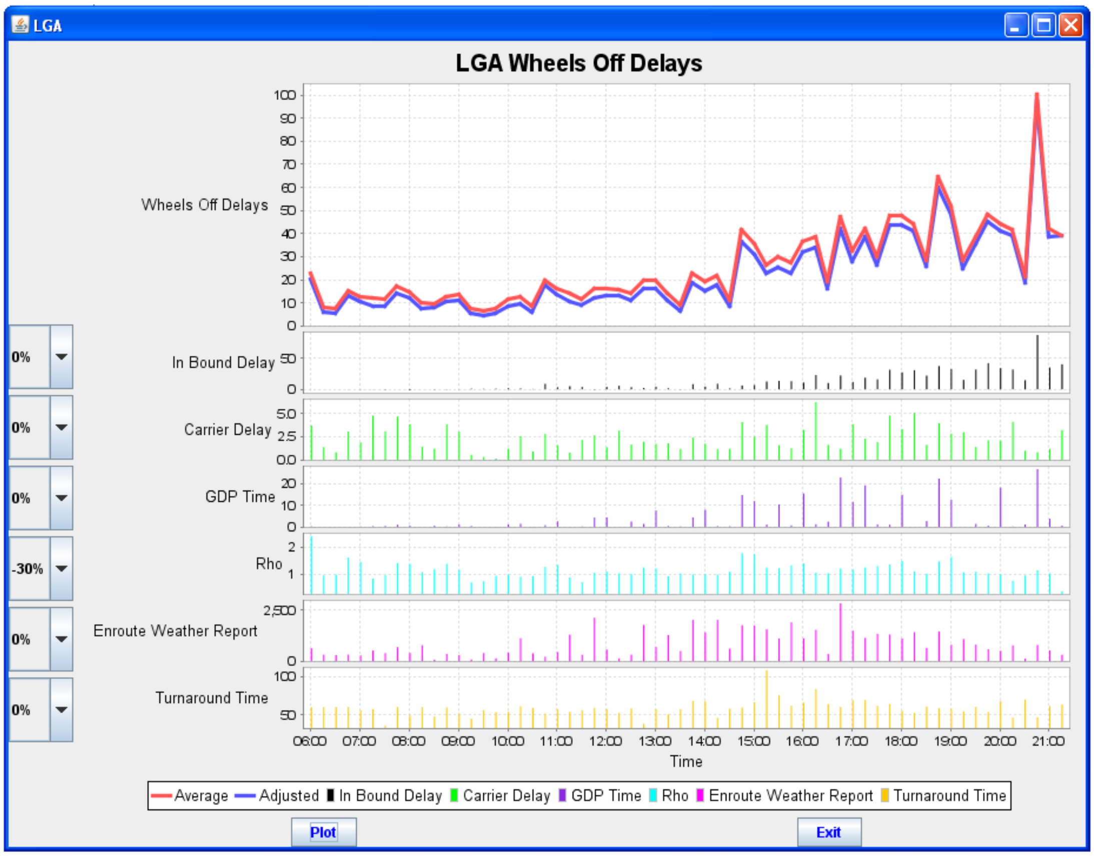


Figure **1.10**: Graphical Display of Results of Implementing Multi-factor Model for Predicting Delay at LGA in 15 Minute Epochs

When collinearity is properly accounted for, the individual influence of factors on response variables can be better estimated. The sensitivity analysis results provide airline and air traffic managers the means to develop and evaluate mitigation strategies.

1.6 Structure of Dissertation

This dissertation is organized as the follows: Chapter 2 provides a review of previous research related to delay causality analysis. These analyses are summarized according to research scope and research methods.

Chapter 3 describes the data process and factor collection.

Chapter 4 provides a brief introduction of the piece-wise linear regression model. A detailed description of the model development and factor selection for final models is represented in this chapter.

Chapter 5 lists examples of final models.

Chapter 6 describes model validation results.

Chapter 7 describes the detailed sensitivity analysis approach and the results.

Chapter 8 describes a case study on 7 airports.

Chapter 9 provides conclusion of results obtained to date and future work.

CHAPTER 2

PREVIOUS ANALYSIS OF DELAYS

2.1 Complex Network Transportation System (CNTS)

A Complex Network Transportation System (CNTS) moves passengers and cargo using multiple modes of transportation and multiple vehicles. There are three internal components of a transportation system: a physical system, management and labor (Sussman 2000). The physical system consists of infrastructure, vehicles, power systems, fuel and control, communication and location systems. The management includes marketing, planning and operations. Four major parts in operating plans are schedule, crew assignment, vehicle distribution and connections. When we treat the transportation system as an interconnected network, the nodes usually represent terminals or stations (for example, airports), and the links are typically guideways (for example, air corridors).

Sussman describes the characteristics of a transportation system as (2000):

1. "Transportation systems are complex, dynamic and internally interconnected, as well as interconnected with other complex dynamic systems (e.g., the environment, the economy)."

2. “They vary in space and time (at different time scales for different components). Service is provided on complex networks. The system is stochastic in nature.”
3. “Subsystems are integrated, closely coupled through feedback loops.”
4. “Human decision-makers with complex decision calculi make choices that shape the transportation system.”
5. “Its behavior is counterintuitive. Developing models that will predict their performance can be very difficult to do.”

Analyzing and measuring the affect of flow of vehicles on transportation networks is a basic element of transportation systems analysis (Sussman 2004).

2.2 Air Transportation System (ATS)

Aviation is one of the critical modes in a national transportation system, which is a large-scale, integrated system. Table **2.1** shows generic components of a CNTS and corresponding components of the Air Transportation System (ATS). In this research, we study OEP-35 airport delays to gain insight into the overall ATS behavior that we can apply to CNTS.

Table **2.1**: Physical System Components of CNTS and ATS

CNTS	**ATS**
Infrastructure: • Guideway • Terminals/Stations	 • Airways, departure and arrival procedure • Airports
Vehicles	Airplanes
Power System	Jet
Fuel	Aircraft gas
Control, Communications & Location Systems	Controller, Sensors, Fleet Management Systems, Air Traffic Central Surveillance, Communication & Navigation Systems.

There are many ways to describe the ATS from different perspectives. Krozel et al. have summarized the behavior of the ATS into a group of descriptive aggregated statistics from aviation related data (mainly 2000) (as shown in Table **2.2**).

Table **2.2**: Example of Statistics Describing the ATS (Krozel et al 2003)

ATS System States	**ATS System Controls**	**ATS System Performance**
Seasonal Trend in Enplanements	Ground Delay Program	Total Gate Departure
Weekly Trends	Cancellations	Airport Departure delay
Average AAR	Ground Stops	Gate delay
IFR vs. VFR	MIT Restrictions	Airborne delay
Airport Visibility		Arrival delay
Runway Configuration Changes		Total taxi in & taxi out delay
		Block delay
		Total delayed operations
		Weather related delays

2.3 Causal Analysis of Delay in ATS

Delay is one of the most pressing problems in the ATS. Because of the major economic and operational impacts of flight delay, it is essential for the Federal Aviation Administration, airlines and other stakeholders to understand the causes of delay and to find ways to reduce delay.

The literature in analysis of delay is categorized into two aspects: 1) delay analysis in ATS by scope (section 2.3.1) and 2) methods for analyzing delays (section 2.3.2).

2.3.1 Delay Analysis in ATS by Scope

Predicting and analyzing the causes of delay have long been important topics of research because of their crucial importance in air traffic management and airline decision making. This problem has been examined from various perspectives, as shown in the Table **2.3**.

Table **2.3**: Summary of Delay Analysis in ATS by Problem Scope

Scope	Method	Strength of Method	Limitations of Method
Macro: NAS	• Independent trend analysis (Krozel et al. 2003) • Regression (Hansen&Hsiao, Hansen&Zhang 2005, Rupp 2005) • Steady state Markov Chain (Boswell&Evans 1997) • Normalization (Evans et al. 2004) • Correlation analysis (Post et al 2002)	• Provides a overall performance of NAS, and general explanation of NAS delay	• Ignores the difference and relationships between airports • Ignores the difference and relationships between delay causes • Not instructive on improving the local airport efficiency.
Meso: Airport	• Classification (Allan et al. 2001) • Cluster analysis (Baden 2005) • Regression (Hansen&Zhang 2005) • Spectra analysis (Welch&Ahmed 2003) • Artificial Neural Network (Dai 2006) • Simulation (Schaefer et al. 2001, Schaefer & Millner 2001, DeArmon 1993)	• Provides causal analysis of single airport • Reveals the correlation among airports • Airport delay patterns reveal their natural grouping	• Ignores the correlation between causal factors. • Simulation model lacks validation • Difficult to introduce time phases • Difficult to separated impact from different airports and the NAS
Micro: Flight Cycle	• Regression (Vigneau 2003) • Analytical (Beatty 1998)	• Includes the system effect into analysis • Provides a dynamic view of delay • Reflects the impacts of flight connectivity	• Absence of airline data • Estimated for specific region and specific airline, not a generic representative for airports in NAS
Micro: Flight segment	• Classification (Allen et al. 2001) • Regression (Callaham et al. 2001, Rupp 2005) • Queuing theory (Idris 2002) • Simulation (Hoffman 2001) • Two-way ANOVA (Willeman 2001)	• Provide detail analysis for each flight segment • The causal analysis is better justified for it can be backed up by physical explanation	• Limited by specific airport or route • The correlations of segments to segments, factors to factors are not explained

The first group of studies uses aggregated delay variables to reveal the relationship between environment variables and ATS aggregated delay. Trend analysis (Krozel et al. 2003), classification (Callaham et al. 2001), and regression analysis (e.g. Hansen and Hansen 2005, Chatterji and Sridhar 2005, Dai 2006) provide a static analysis of the ATS at a single point in time. Boswell and Evans (1997) developed a steady state Markov Chain model to examine temporal dynamics of delays and to study the correlation of the current ATS performance to the previous time period's ATS performance.

From an air traffic administration perspective, at any time period during a day, the understanding and prediction of all flights' on-time performance is not only necessary to even out the demand on the limited capacity, but essential to improve the local airport's and ATS' efficiency. Research from an airport administration perspective which ignores the difference among airlines can be categorized into four groups: (1) study a local airport while ignoring its correlation with other airports in NAS (Allan et al 2001); (2) study the impact of delays of outbound flights from a specific airport to other airports in NAS (e.g. Schaefer and Millner 2001); (3) study the impact delays of inbound flights from airports in the NAS to a specific airport (DeArmon 1993); (4) study the relationship of aggregated delays of inbound flights from the entire NAS and outbound flights to the entire NAS (Baden et al. 2005).

The major difficulty in meso-level airport delay analysis is determining how to introduce different time phases into the analysis. ATS is a connected dynamic network.

Due to the difference in airport configurations and the geographic distance, the transaction time and the strength of impacts from an event at one airport to other airports depend on the distance between airports and the airport's ability to absorb the impacts. The questions can be, for example, "Why does Miami airport have high delays when the weather is perfect and the airport is not congested?" and "Where did the delayed flights to the Miami airport come from?"

To tackle these problems, researchers go to a lower level, the flight level. Three types of downstream impacts from delay propagation have been summarized by Boswell and Evens (1997): downstream delay, flight cancellations and missed connections. They pointed out the downstream impacts were considered to be the "major and sometimes dominant factor in assessing the total costs of air traffic delay". Two major concerns an airline has about their delayed flights are "when disrupting events happen, how well will the affected airframe complete its whole day schedule (Vigneau 2003)". And "how will this airframe disturb the airlines total schedule (Beatty et al 1998)?" Few studies have been done on airline flight cycle because of the absence of airline data.

A summary of the research on the factors influencing each flight segment is reported in Table **2.4**.

Table **2.4**: Causal Analysis of Flight Segment Delay

Paper	Segment	Method	Factors
Hoffman 2001	Arrival and Departure	TAAM simulation	Levels of traffic
Wang et al. 2003	Turn Around	Analytical model	Slack and flight time allowance
Idris et al. 2002	Taxi Out	Queuing model	Runway configuration, terminal, weather, downstream restriction, departure demand, queue size
Welch and Ahmed 2003	Arrival, Airborne	Spectra analysis	airport throughput
Willemain 2001	Airborne	Two way ANOVA	Airspace of origin, destination and end route

These flight segment analyses provide detailed analysis of the impacts of several causal factors. These causal analyses are intuitive for they can be backed up by physical explanation, but they are limited by the problem scope. The analysis is only for a specific airport or route. After obtaining the flight information of routes and airports, there are many questions to be answered, such as how do airports influence other airports; how are airports influenced by other airports; how does an individual airport influence NAS; and how does NAS influence individual airports.

2.3.2 Delay Analysis in ATS by Methods

The methods for analyzing delays can be summarized into five categories: regression and related methods, time series analysis, Bayesian Networks model analysis, cluster and classification analysis, and simulation. The section on regression analysis includes methods for using observations to predict or explain delay. The methods include linear regression, neural networks and related methods. In time series analysis, the trend analysis, spectral analysis and Markov Chain analysis are introduced. Linear classification and cluster analysis are described in the classification section. The appropriate problems, assumptions, strengths, and weaknesses of each method are described in detail in the following sections.

In this document, we use the capital letters to represent random variables and a letter in bold face to represent a vector or a matrix.

2.3.2.1 Direct Measure

Allan et al (2001 a) described a "direct" method to determine the achieved delay reduction benefits with the combination Integrated Terminal Weather System and Terminal Convective Weather Forecast (ITWS/TCWF). This direct measurement "compares the delays in a baseline time period when ITWS/TCWF were not in use to a subsequent time period in which ITWS/TCWF were in use." But as these authors have pointed out the reduction in delays can be a result of many factors such as severity of weather, duration of weather, traffic changes, air traffic procedures, etc. Hence, the explanation of result requires sophisticated analysis of causality in order to distinguish which elements of the systems account for the reduction.

Another analysis of delay at Newark International Airport was conducted by Allan et al to determine causes of aviation delay. The delays were categorized into groups defined by causes. The proportion of each group was computed as the consequence from the resident cause. This kind of method is widely used in transportation statistical reports (DOT) but there are two drawbacks of it. First, it cannot be used to predict delay. Second, the assumption that these delay causes are mutually exclusive is not valid all the time.

2.3.2.2 Regression Analysis

In an univariate regression analysis, there is one numerical response variable (dependent variable) *Y*, and one or more predictor variables (explanatory variables or independent variables) ***X***. A goal of regression analysis is to predict the unknown value of the response variable associated with a given set of known predictor values. Additionally, one may desire to determine the mean and variance of the response variable conditioned on the predictor variable values. Another goal is to represent the independent contributions of each predictor variable to the prediction of the response variable using the regression coefficients. In other words, regression provides a way to determine how the variable X_i influences *Y* after controlling for other independent variables.

Table **2.5** summarizes research that used regression analysis for causal analysis of delay in NAS. The first column in Table **2.5** describes the problem to be solved. Researchers attempted to distinguish the contributions of several critical factors, called predictor variables, to delay; the predictor variables are listed in column two. The researchers and their research approaches are listed in the third and fourth columns, respectively.

Table **2.5**: Summary of Research Using Regression Analysis

Response variable	Predictor variables	Author	Methods
Probability of flight on time, delay or cancellation at LGA,	Economic (e.g. revenue, load factors), Route Competition (e.g. monopoly), Airport Competition (e.g. concentration at origination, hub destination), Logistical (e.g. slot origination, distance, hours until next flights), Weather (e.g. rain, minimum temperature, frozen)	Rupp 2005	Nested logit model
Probability of flight delay and cancellation at NAS			
AAR	scheduled arrivals, visibility, wind speed, interaction of the visibility and operational condition, time and season	Hansen & Zhang 2005	GARCH
LGA daily avg. arrival delay,	NAS daily average arrival delay	Hansen & Zhang 2005	Two-stage least squares model with GARCH model
NAS daily avg.e arrival delay	LGA delay		
	Exogenous variables: derived queuing delay, adverse weather, EDCT holding and total flight operations		
Avg. daily delay	Arrival queuing, convective weather, terminal weather conditions, season, secular effects.	Hansen & Hsiao 2005	GARCH
Weather Weighted Traffic Count (surrogate for system delay)	Expected traffic demand, Weather Impacted Traffic Index (WITI), IMC, wind speed	Chatterji & Sridhar 2005	Ordinary Least Square (OLS)
Arrival delay	Departure delay, origin	Vigneau 2003	Recursive OLS
Departure delay	Arrival delay, destination, load on departure, station stoptime		
	day of week, hour, destination, capacity, load on arrival		
NAS impact (delay>30min, cancellation, diversion)	En route WITI, IMC< wind speed	Callaham 2001	OLS
	Weather day-type, traffic day-type, season		
Taxi out time	Runway configuration, airline, VFR/IFR, downstream restrictions, departure demand, queue size	Idirs et al. 2002	OLS

2.3.2.2.1 Ordinary Least Square (OLS) Regression

Appropriate problems:

OLS regression models relationships between the response variable and predictors. It is usually applied to observed sample or experimental data on a response variable, and tries to explain the behavior of that variable in the form of an algebraic equation that involves other variables (predictors) that describe experimental conditions.

***Assumptions**:*

Suppose there is a response variable *Y* and predictor variables ***X.*** The objective of regression is to develop a statistical model to predict *y* from ***x***. The simple linear regression model represents the relationships between *Y* and ***X*** (as shown in Equation **2.1**).

$$y_i = \alpha + \boldsymbol{\beta}^T \boldsymbol{x}_i + \varepsilon_i \qquad \textbf{2.1}$$

In Equation **2.1,** y_i are particular observed values of the target variable *Y*, $\mathbf{x}_i$ are the corresponding observed values of the predictor variables, and ε_i are unobservable random errors, or "noise". α and $\boldsymbol{\beta}$ are a set of unknown parameters that are estimated based on the observations of *Y* and ***X***. The estimates of the parameters are denoted by $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$. $\hat{\alpha} + \hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i$ yields estimated values of y_i, denoted by $\hat{y}_i$, as a function of $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$. Higher order polynomial regression allows for a nonlinear relationship between response and predictors. It is still called linear regression since the model is linear in the parameters.

There are 3 classical regression assumptions (Hamilton 1994): 1) Independent variables $\boldsymbol{X}$ are a vector of deterministic variables and measured without error, 2) ε_i are independent and identically distributed (i.i.d.) with mean 0 and variance σ^2, and 3) the distribution of ε_i is Gaussian.

Linear regression models assume any nonlinear relationship between predictors and *Y* can be transformed into linear form by applying transformations to the predictor variables (when this assumption is not met, nonlinear regression methods may be used). The least square estimates of α and $\boldsymbol{\beta}$, which minimize the sum of the squares of residuals $\sum(y_i - \hat{y}_i)^2$, are optimal when errors are independent, normal, with zero mean, and homosecedastic.

Strengths and weaknesses:

The strength with regression analysis is the ability to use a single regression equation that captures effects of all predictors simultaneously. Despite the variability in the data, the regression analysis can isolate the effect of each variable from the multitude of variables in the study and provide an explicit quantitative estimate of the impact of each predictor. The estimated coefficient, $\hat{\beta}_j$, for the *j*th predictor gives the estimated change in $E\left(Y_i \mid \mathbf{x}_i\right)$ given a unit change in the *j*th predictor.

The problems with regression methods come from two sources: data problems and model problems (Freund et al 2006). Specifically, the weaknesses of OLS regression are:

1) A major type of data problem involves observations which can greatly influence regression model estimates. Influential observations include outliers (extreme observations in the response variable) and leverage points (extreme values of predictors). Many statistical methods have been developed to detect outliers but the remedy for the effects of outliers is not purely a statistical problem. The outliers may be either recording errors or observations that are subject to a factor which is not included in the model. For this reason, whether to keep, correct, or remove extreme observations needs careful investigation.

2) A problematic regression model can be either overspecified or underspecified. Overspecified models include too many explanatory variables. The existence of collinearity among predictor variables causes the estimated regression coefficients to have inflated standard errors and also makes it difficult to distinguish their individual influence on the response variable. Underspecified models miss important predictors and/or incorrectly specify the relationships, for example, representing a non-linear relationship as linear. So, a tradeoff between bias and variance has to be made since estimating a large number of unknown parameters may overfit the model and increase the variance of the estimation, and deletion of a predictor may create bias.

3) The assumption of independent, identical, normally distributed error terms is rarely met in practice, especially in air transportation systems. Heteroscedasticity, unequal variance, can be the result of a data problem, or a model problem, or both. Various methods have been applied to compensate for this problem. Generalized Autoregressive Conditional Heteroskedasticity (GARCH) in Hansen's paper (Hansen & Zhang, Hansen & Hsiao 2005) and White (1980) robust standard errors in Rupp's paper (Rupp 2005) both address the existence of heteroscedasticity in error terms. The GARCH models are explained in the time series method section. The model from White did not report a good fit (Pseudo $R^2<0.1$).

4) The coefficients of predictor variables reflect the unique contribution of each predictor. However, if a predictor has joint contribution with other factors, the coefficient of their interaction term does not reflect the contribution of any particular predictor

. 5) A regression model, by itself, does not imply cause and effect. Regression is a way to measure correlation between response variable and predictors. Correlation implies causation only when the data from which the correlation was computed were obtained from controlled experiments. Controlled experiments control for extraneous variables which might confound the results. Causal interpretation of regression results for non experimental data requires careful justification based on knowledge of the domain.

In NAS delay analysis, two additional kinds of models, two-way ANOVA and Artificial Neural Networks, can also be categorized as regression models.

2.3.2.2.2 Two-way Analysis of Variance (Two-way ANOVA)

Appropriate problems:

Analysis of variance (ANOVA) tests for significant differences between group means through the portioning of Sums of Squared (SS) Deviations. It has been widely conducted on data collected in experimental studies to infer causality. In two-way ANOVA, there are two independent variables or factors (e.g. origin and destination in Willemain 2001). The hypotheses to be tested by two-way ANOVA according to Willemain's paper are:

1. The population means of the first factor are equal, i.e., the different flight origin airspace does not have different impact on daily average airborne delay;
2. The population means of the second factor are equal, i.e., the different flight destination airspace does not have different impact on daily average airborne delay;
3. There is no interaction between the two factors, i.e., the flight destination airspace's impact on airborne delay does not depend on the origin airspace.

***Assumptions*:**

The ANOVA test is, practically, equivalent to a regression analysis (Draper and Smith 1981). The ANOVA model in Draper and Smith can be represented mathematically as a multiple regression with dummy coded predictors, as shown in Equation **2.2**.

$$Y_{ijk} = \mu + R_i + C_j + (RC)_{ij} + \varepsilon_{ijk} \quad \textbf{2.2}$$

where, μ = overall mean,

R_i = effect of *i*th row factor,

C_j = effect of *j*th column factor,

$(RC)_{ij}$ = interaction effect of row *i* and column *j*,

ε_{ijk} = random error within the (I,j)th cell for k observations assumed to be distributed as $N(0, \sigma^2)$. Errors are assumed to be pairwise uncorrelated.

Jones summarized the assumptions of the two-way ANOVA *F* test as (Jones 1996):

1. "The populations from which the samples were obtained must be normally or approximately normally distributed.
2. The samples must be independent.
3. The variances of the populations must be equal.
4. The groups must have the same sample size."

Strengths and weaknesses:

Two-way ANOVA analysis provides a mathematical way to test the effects of two factors. If the data was collected through controlled experiments, the results from ANOVA tests can be used to infer causality. But there are some noteworthy theoretical problems of the ANOVA test as summarized by Notthcott (2006). First, by selecting different range of factors or different choice of population, the values ANOVA yields for causal efficacies can be completely different. The ANOVA cannot reliably predict the outcomes of interventions when the input is outside the range of a generating sample. It can be misleading to extrapolate ANOVA results uncritically (Lewontin 1974). Second, ANOVA assigns causal efficacies only at the group level rather than singleton level. For example, origin airspace is a group level factor. It includes airspace of many airports which are at singleton level. The ANOVA *F* result, by itself, does not tell which components of the origin airspace cause more airborne delay.

In practice, the assumptions of an ANOVA test are rarely met when data is not gathered from a carefully designed experiment. For example, the condition of airspace above Honolulu Airport (HNL) may be completely different from and not related to the airspace above LaGuardia Airport (LGA), but the airspace above Newark International Airport (EWR) is not likely to be much different from its neighboring airport LGA. The sample gathered about HNL airspace may be independent from the sample of LGA. However, the sample of EWR is not independent from the sample of LGA.

2.3.2.2.3 Artificial Neural Network (ANN):

"An ANN is a model composed of several highly interconnected computational units, called neurons or nodes" (Detienne et al. 2001). These nodes are categorized into input layer (independent variables), output layer (response variables) and hidden layer (computation procedures) according to their position and function in the network.

Appropriate problems:

ANN can be viewed as a nonlinear regression method. ANN represents the relationship between y and x as:

$$E(y) = f(\boldsymbol{x}, \boldsymbol{\beta}), \quad \textbf{2.3}$$

where f is some function of the observations $\boldsymbol{x}$ and parameters $\boldsymbol{\beta}$.

Since there is no closed form solution to estimate $\boldsymbol{\beta}$, iterative methods are used to solve for a local minimum of some measure of residuals, usually squared errors.

Assumptions:

ANN can be applied with very weak assumptions. When the relationship (either linear or non-linear) between input variables and output variables are estimated by ANN, it is generally assumed that the estimation errors from the learned model are independent and identically distributed with zero mean.

Strengths and weaknesses:

The most important advantage of ANNs is that they can solve complex problems which do not have a closed form equation for estimating the regression parameters or for which an algorithmic solution is too complex to be found (Stergiou and Siganos 1996). Dai (2006) developed an artificial neural network model to estimate individual flight departure delay for real time air traffic flow management. A network with 70 nodes in hidden layers outperformed linear and non-linear regression method in Dai’s research.

A neural network is a “black box” model that predicts departure delay from the input factors. The parameters of a neural network model are not easily interpretable, and thus it is difficult to use a neural network model to gain understanding of how the factors interact to cause delay. The iterative learning process of ANN requires long processing time as the size of the problem expands (Stergiou and Siganos 1996). Another weakness of neural networks is the lack of defining rules to construct a network. The choice of learning algorithm and model requires relatively good understanding of the underlying theory. The number of neurons per layer and number of layers are usually selected in the learning process after several trials. Also, there are no standard statistical tests for ANN models. When the relationship is nearly linear and error s are approximately normal, a correctly specified multiple linear regression is superior to ANN (Warner and Manavendra 1996).

2.3.2.3 Time Series Analysis

Time series analysis examines data that has been observed at different points in time. In most cases, these observations violate the assumption of the traditional regression method that observations are independent. The systematic approach of drawing inferences from these time-correlated series is referred to as time series analysis.

A general approach to modeling the dependence over time consists of the following steps (Brockwell and David 1996):

- "Plot the series and examine the main features of the graph, checking if there is a trend, a seasonal component, any apparent sharp changes in behavior and any outlying observations.
- Remove the trend and seasonal components to get stationary residuals.
- Choose a model to fit the residuals.
- Forecasting will be achieved by forecasting the residuals and then inverting the transformations described above to arrive at forecasts of the original series. "

In the air transportation field, researchers usually look at the linear component of yearly, seasonal and monthly trends. Hansen et al. conducted a series of studies on the trend of delays and cancellations, such as the delay trend before and after the enactment of Aviation Investment and Reform Act for the 21st Century (AIR21) at LGA. In these studies, the residuals were fit against the past values of the series by Generalized Autoregressive Conditional Heteroskedasticity (GARCH) (Hansen and Zhang 2005, Hansen and Hisao 2005). Beside the complex GARCH model, a simple time domain approach, trend analysis, and frequency domain approach, spectral analysis, were conducted in the NAS delay analysis field.

2.3.2.3.1 Trend Analysis

Appropriate problem:

Trend analysis has been widely used to visualize the shape of one quantitative dependent variable related to a quantitative independent variable, time t. The linear component of trend t vs. Y is used to evaluate whether there is an overall increase or decrease of Y as the independent variable t increases. Rupp looked at the yearly trend of NAS delay, cancellation and on time arrivals (Rupp 2005); Krozel et al. plotted the seasonal trends of NAS states, NAS control and NAS performance (Krozel et al. 2003). These analyses applied the first step of time series modeling, plotted the series and examined the main features of the graph, checking if there is a trend, a seasonal component.

Assumptions:

Simple trend analysis assumes time to be the dominant independent variable to predict changes in the response variable.

Strengths and weaknesses:

The trend plot provides direct visualization of change direction and degree of response variable. If the residuals are independent and identically distributed (i.i.d.), the independent variable, time, is the dominant independent variable to the changes of response variable. The tendency in practice of plotting the mean of response variable vs. time masks the variance of the response variable at each time period. That variance may also relate to time. Furthermore, the trend plot ignores the influence from other variables.

So, when there are many interacting variables, trend analysis provides an incomplete picture of delay.

2.3.2.3.2 Spectral Analysis

Appropriate problem:

Spectral analysis uses frequency domain analysis to identify superimposed oscillations which contribute to variations in series of observations. Spectral analysis is a frequency domain approach in time series analysis.

Spectral analysis is applied to a signal, $y = f(t)$, where t typically corresponds to time and y corresponds to the amplitude of the signal. Spectral analysis breaks the signal down into frequency components with suitably chosen amplitudes and phases. These amplitudes and phases arise from the Fourier Transform of the signal. An example is given in Equation **2.4**.

$$y = \alpha_1 \sin(\theta_1 t) + \alpha_2 \sin(\theta_2 t + c_2) + \varepsilon \tag{2.4}$$

Where, t is time

α_i is the amplitude at a frequency

θ_i is the a particular frequency

c_i is phase which measures the offset in time between the waves

ε is random error

The problem of spectral analysis is to estimate the discrete set of frequencies, amplitudes and phases from a sample of the signal at given times. Welch and Ahmed applied spectral analysis on airport throughput and delay in order to separate the effect of a new Phoenix runway from the effect of background noise, such as en route congestion, weather, and flow control actions (Welch and Ahmed 2003). The distribution of occurrences of periods of observed arrival throughputs at an airport is called the occurrence spectrum for the airport, and the related plots of delay versus throughput are called the delay spectra. Different relationships among arrival delay, queuing delay and arrival throughput comprise different airport throughput spectrum. Welch and Ahmed tried to distinguish the benefit of congestion management initiative from the different patterns of throughput and occurrence spectra for U.S. airports operating near capacity. They found that the hub and non-hub airports can be discriminated by the different patterns of arrival delay and queuing delay at the highest end of throughput.

Assumptions:

In spectral analysis, the regularity of a series, a repeated pattern with a regular time interval, is composed of a superposition of periodic variations of the underlying phenomenon that produce the series (Shumway and Stoffer 2000).

Strengths and weaknesses:

Spectral analysis has the advantage of being a simple measurement capable of evaluation of the contributions of the periodic components. The spectrums provide valuable insights to the problem. In a CNTS, there are typically many causal factors whose influence must be considered. Standard spectral analysis techniques do not consider how the signal depends on independent variables other than time.

2.3.2.3.3 Markov Chain Modeling

Appropriate problem:

A Markov chain is a special type of discrete-time stochastic process (Winston 1993). Since Markov chain modeling studies how a random variable changes over time, it was grouped into the time series analysis section in this dissertation. Boswell and Evans developed a Markov Chain model based on the data from FAA database to predict downstream delays in NAS and likelihood of flight cancellation as a function of the initial delay in NAS (Boswell and Evans 1997).

Assumptions:

A Markov chain model assumes the system can be in one of a discrete set of states. A Markov chain represents a process starting in one of these states and moving successively from one state to another assuming the probabilities p_{ij} (moving from state i to next state j) only depends on its current state i and not on its state prior to the current state.

In the Boswell and Evans model, the arrival delay at the next leg, $(i+1)th$ leg, is summation of downstream portion of arrival delay at previous leg and operational delays incurred in the current leg. The distribution function of arrival delay at the second leg is the convolution of the distributions of the two summands, operational delay and the previous leg carryover delay, under the assumption that they are statistically independent random variables.

To calculate a general-purpose multiplier for downstream delay, another assumption was made to assume the direct delays are uniformly distributed among flight legs, i.e. the flight delay may occur on the 1^{st} leg, 2^{nd} leg,..., 6^{th} leg with the equal possibility.

Strengths and weaknesses:

A Markov chain model is a very simple, direct representation of a dynamic system whose components are connected sequentially.

In reality, many systems are too complex to represent as stationary first-order Markov chains. Transitions may depend on states prior to the current state. The assumption of homogeneous transition behavior is often violated because different subgroups may have different transition propensities.

In the Boswell and Evans paper, the calculated downstream multiplier is a constant ratio, 80%, between an initial delay and its total downstream impacts from flights at all sites and in all weather and traffic conditions. Given the various weather conditions and traffic patterns, adaptive nature of air traffic control and airline operation, and airport state (hub, or non-hub), it may not be realistic to assume the same transition matrix for all airports, weather and traffic conditions.

2.3.2.4 Discrete Bayesian Network Analysis

A Bayesian network (BN) is a directed acyclic graph, in which each node denotes a random variable, and each arc denotes a direct dependence between variables. The BN model structure (nodes and arcs) encodes conditional dependence relationships between the random variables. Each random variable is associated with a set of local probability distributions (parameters in the Conditional Probability Tables). Probability information in a Bayesian network is specified via these local distributions. A root node in a BN model represents a random variable and its associated probability distribution. A non-root node has an associated random variable and a conditional distribution for its random variable given the values of the parent random variable(s).

A discrete BN model contains only categorical variables and discretized continuous variables with a finite number of possible states. Currently, most standard BN software packages are limited to discrete variables.

Appropriate problems:

Bayesian Networks provide a powerful and expressive language for representing a complex phenomenon involving uncertainty and probabilistic reasoning. Rather than representing just one variable conditioned on its influencing factors as in a regression model, a BN model represents a joint probability distribution for all its nodes. BN models have gained increasing popularity for reasoning and decision-making. Through a well defined inference algorithm, a BN model can infer the probabilities of variables which have not yet been observed based on the observations of known variables.

Assumptions:

David Danks summarized two assumptions for a Bayes Net formalism: Markov assumption and Faithfulness assumption. The Markov assumption, "X is (probabilistically) independent of its (graphical) non-descendants conditional on its (graphical) parents", uses the absence of an arc to imply conditional independence. The Faithfulness assumption, "The (probabilistic) effects of (graphical) paths never exactly offset", derives the absence of arc from conditional independence (Danks 2003).

The BN model structure and parameters are either learned from data or constructed via expert judgment. It is common to link the cause node to the result node if the model represents a cause and effect relationship, or to link the nodes time order. Orienting links from cause to effect generally results in sparser networks (Pearl 2000).

Strengths and weaknesses:

Bayesian networks have become an increasingly important tool for investigating interdependence among multiple factors in complex systems. Bayesian networks have unique strengths both for inference and for visualization. When Bayesian networks are combined with traditional statistical methods, conditional independence can be exploited to provide more accurate estimation and therefore more precise prediction (Xu et al. 2005).

For large problems, exact inference in discrete Bayesian Networks is intractable. More sophisticated non-parametric density estimation methods and approximate inference algorithms tailored to continuous distributions are needed to model complex system and reduce discretization error.

2.3.2.5 Classification and Clustering

Classification and clustering methods are related to partitioning data into groups. Classification labels data according to the rules learned from the data with known group. Clustering partitions data into clusters.

2.3.2.5.1 Classification

Appropriate problems:

Typically, classification is about labeling groups of things according to their shared common characteristics. There is a response variable as well as one or more predictor variables in classification as in regression. The classification rules (classifier) are learned from known labeled cases, where the values of both predictor variables and response variable are known. After a classifier has been learned, it is then applied to unlabeled cases to allocate them to previously defined groups.

Suppose that we have K mutually exclusive, exhaustive groups, $Class_1$, $Class_2$, …$Class_k$, and a p- dimensional vector, $\mathbf{Feature}$. The prior probability $P(Class_i)$ and conditional distribution of the $\mathbf{Feature}$ given $Class_i$, $P(\mathbf{Feature} \mid Class_i)$, have been estimated from a sample of labeled cases, $\mathbf{Feature}_1, \cdots, \mathbf{Feature}_n$, called the training sample. Classification techniques are developed for solving the following problems. 1) Estimating the class-conditional probability $P(\mathbf{Feature} \mid Class_i)$, if all variables are discrete; 2) Considering classification as a prediction problem, where the goal is to estimate a function for $P(Class_i \mid \mathbf{Feature}_{new})$.

Assumptions:

The basic statistical assumption underlying classification methods is that observations are independent and identically distributed within each class, and the classes are mutually exclusive.

Strengths and weaknesses:

Various classification algorithms have been developed to construct simple or complex linear or non-linear classifiers. Classification methods are generally more appropriate than regression methods if the response variable is categorical. If the response variable being predicted is numerical, regression methods are more appropriate.

Linear Classifier:

Appropriate problems:

In the existing literature, a classification technique applied to analyze NAS delay is linear classifiers (Allan et al. 2001, Chatterji and Sridhar 2005). In these analyses, the conditional distribution of the Features given these delay has or has not occurred, $P(\mathbf{Feature} \mid Delay)$, are estimated from the historical data, where *Delay* is a categorical variable with states Yes or No. A typical statement is "41% arrival delay occurred on days characterized by convective weather". In these studies, the goal for creating a classification rule is to gain a better understanding of a certain phenomenon but not to predict the probabilities of delay given the weather conditions.

Assumptions:

A common linear classification method, linear discriminant analysis (LDA), assumes Gaussian conditional density models and the decision boundaries to separate groups are linear. It is further assumed that the normal distributions for the classes all have the same covariance matrix and only distribution means are different.

Strengths and weaknesses:

The LDA classifiers are simple and straightforward. For a two-class problem, the operation of a linear classifier splits a high-dimensional input space with a hyperplane. One side of the hyperplane is “yes” and the other is “no”. But if the optimal decision boundary is not close to linear, then linear classification can perform poorly. In some cases, expanding and/or transforming the set of predictors can improve performance.

2.3.2.5.2 Cluster Analysis:

Appropriate problems:

Cluster analysis is used to identify natural clusters in a set of cases. The cases within a cluster are more similar to each other than they are to cases in other clusters.

The application of cluster analysis in NAS delay problem focuses on weather normalization (Callaham et al. 2001) and airport categorization (Baden et al. 2006). The purpose of weather normalization done by Callaham et al. is to determine the discrete weather day-types with respect to severe en route weather together with the determination of traffic day-types, so that a fair comparison of NAS performance over different time intervals can be conducted given the fixed impacts from weather and traffic.

Baden et al. conducted cluster analysis on the delay influence at local airport (relationship between inbound and outbound delays) and downstream airports (accumulated amount of delay transmitted by airframes). Two analyses yielded three different clusters of airports, but the airports geographically located closer are usually in the same cluster, and the hub airports tend to be in the different cluster from the non-hub airports.

***Assumptions**:*

Cluster analysis is conducted without prior assumptions on original data distribution and range.

Strengths and weaknesses:

The strength of cluster analysis lies in its capacity to discover the natural categories of high dimensional data; it has minimal requirements of domain knowledge and is capable of handling noise and outliers. There are various ways to build clusters, and there can be subjectivity in the choice of method. Applying different clustering methods, such as Ward's Minimum Variance, Average Linkage, and two-stage Density Estimation, on the same set of data yields different clusters. Hence, there is not a single correct way to separate data into clusters.

None of the papers cited in this section explained what kind of measurement or rule was applied. It would give the reader a better understanding of their data structure if they had provided that information. Furthermore, cluster analysis is a useful starting point for other research purposes (see for example Tan et al. 2005).

2.3.2.6 Simulation Methods

Appropriate problems:

Simulation is a way to predict outcomes of a process for which an executable model has been developed (e.g. DeArmon 1993, Schaefer et al. 2001, Hoffman 2001). Regression analysis, time series analysis, and classification are all statistical methods, which use data to develop a model of the process that generates data. Simulation methods use the developed models to predict the phenomenon of that process given different settings.

Simulation has been widely applied in the aviation area because it is too costly and intractable, in many cases, to do experiments on the system itself or to solve a model analytically. Table **2.6** listed 11 simulation models for capacity and delay as selected by Odoni et al. (1997).

Table **2.6**: Summary of Capacity and Delay Models

	Scope of Model			
Level of Detail (type of study)	Aprons and taxiways	Runways and final approaches	Terminal area airspace	En route Airspace
Macroscopic (Policy analysis, cost-benefit studies)		LMI Runway Capacity Model* FAA Airfield Capacity Model* DELAYS* AND*		ASIM SDAT* DORATASK
Mesoscopic (Traffic flow analysis, cost benefit analysis)		NASPAC TMAC FLOWSIM ASCENT		
Microscopic (Detailed analysis and preliminary design)	TAAM SIMMOD			
Same	The Airport Machine		RAMS	

Several simulation models have been developed after Odoni et al published the paper about simulation models. One of them is Future Air Traffic Management Concepts Evaluation Tool (FACET) developed by NASA. This tool models system-wide en route airspace operations over the contiguous United States (Bilimoria and Sridhar 2000). Airport and Control Center Simulator (ACCES) is another simulation model developed by NASA, which integrates the simulation models of airport operation and en route operation.

Assumptions:

A simulation model is developed based on the understanding of how the system works. If the model is a faithful representation of the real-world system, then what can be learned from the model will accurately represent what would have been learned about the system by direct manipulation.

Strengths and weaknesses:

The main advantage of a simulation model is its ability to deal with a very complicated system. The weaknesses also come from the complexity of the problem to be solved. Real systems are usually affected by uncontrollable and random inputs. There are many uncertainties in the system. However, most simulation models for large scale complex system tend to be oversimplified.

Many of the models listed in Table **2.6** were integrated from queuing models of individual flights. But the near-capacity operation and highly adaptive nature of the system make the standard assumptions for queuing theory inapplicable. Many times, decision makers take specific actions that are different from the simulation. The actual system may include important factors that are not represented in the simulation. Hence, the real system reaction may be very different from the outputs of simulation.

2.4 Summary

Existing research has provided the basis for delay analysis, but for the purpose of analyzing causes of delay in the ATS, there are some noteworthy limitations:

- Detailed airline perspective models are too detailed to represent the system performance.
- Analytical models of the aggregated system are too simplified to reflect the interrelationships between airports and the whole system.
- No analysis has been done for individual airport delays at 15 minute increments for a network including a majority of the major airports.
- Data mining results on the airport network of system cannot predict the response of the system to changes in the network.
- Simulation models are based on many assumptions. These assumptions and model outputs need to be validated and need to represent the inherent stochastic and adaptive nature of the system.

The purpose of this research is to apply proper data analysis technique to identify the important factors and to measure the degrees of their impact on delays at 34 OEP airports, so that an accurate prediction of airport delay at a practical level of detail can be achieved. Regression analysis, as a theoretically sound and well-developed technology, is a sensible way to accomplish this purpose. Regression analysis can not only predict the unknown value of the response variable associated with a given set of known predictor values, the variable coefficients in a linear model can also represent the independent contributions of each predictor variable to the prediction of the response variable as well.

Section 2.3.2.2.1 described two sources of problematic regression models: data problems and model problems (Freund et al 2006). The next section, Chapter 3, describes methods and processes conducted in this research to minimize data problems and model inadequacy, i.e., removing influential observations and collecting as many factors as possible to avoid missing important predictors. Chapter 4 explains the method applied to model non-linearity and balance model overspecification and inadequacy.

CHAPTER 3

DATA PROCESSING

Section 3.1 introduces the data source and the development of the Center of Air Transportation Systems Research Center (CATSR) Delay database. The CATSR Delay database includes a Flight Delay module and an Airport Delay module. The Airport Delay database is built on the top of the Flight Delay database. Section 3.2 describes factors studied in the literature and explains how factors were calculated in this research.

3.1 Database Development

3.1.1 Data Sources

The data used in this research comes from the Aviation System Performance Metrics (ASPM) and Enhanced Traffic Management System Counts (ETMSC) of the FAA database, Airline On-Time Performance Data from Bureau of Transportation Statistics (BTS), and National Convective Weather Detection (NCWD) databases.

FAA

The data collected from ASPM are Individual Flights Report and Quarter Hour Airport Report. The Individual Flight database includes flight data, ground and flight movement times. Arrival delay and departure delay are obtained by comparing flight times to air carrier schedules from the Official Airline Guide (OAG) and carrier reservation systems. In the ASPM database, early departures and arrivals are assigned zero delay. In order to incorporate more detailed information about each component of delay for this research, negative delay values were computed for flights that arrived earlier than scheduled.

The ASPM Quarter Hour Airport Report records the counts of airport operations categorized in various groups and average delays in each 15-minute epoch, and the airport condition information and airport efficiency information. The airport condition information includes airport meteorological conditions flag, airport supplied runway configuration, ceiling, visibility, temperature, wind angle, and wind speed.

ETMSC provides the aircraft equipment of flights between city pairs. This database is the source for aircraft type and weight in the CATSR Delay database.

BTS

The Airline On-Time Performance Database from BTS contains data pertaining to US certified air carriers that account for at least one percent of domestic scheduled passenger revenues (BTS 1). Hence, there are fewer flights recorded in this database of BTS than in ASPM. Both ASPM and BTS databases have data fields for flight number and aircraft tail number (unique identifier of aircraft). The flight numbers of some aircraft were erroneously recorded as tail numbers in the ASPM; these incorrect tail numbers were corrected by the records in the BTS, based on flight arrival, departure time and carrier information.

BTS does not have detailed flight segment information, but the information on a cancelled flight, the reason for cancellation, and diverted flights provide a good supplement to the ASPM. Another type of unique information recorded in BTS is the identification of Carrier Delay and Security Delay. They are explained in detail in section 3.2.

Both ASPM and BTS count a flight as "on time" if it operated less than 15 minutes **later** than the scheduled time shown in the carriers' Computerized Reservations Systems (CRS). In the database developed in this research, all flight delays are recalculated to include early flight information, i.e., negative delays.

The flights in CATSR Airport Delay database are the joint set of BTS and ASPM; hence, they are about the flights recorded in BTS.

NCWD

One type of data collected from NCWD is the hourly weather condition at major airports in categorical format. When more than one type of weather exists at the same time, the worst condition for airport operation will be recorded. The other type of data in NCWD is the hourly total number of severe weather reports along each flow. This data was treated as a surrogate for the en-route weather condition but in numerical format. Zero means no severe weather on that flow. A larger number of reports corresponds to worse weather conditions. Together with the ASPM Quarter Hour Airport Report, these two databases are the sources of weather related factors in our developed database.

3.1.2 Flight Delay Database

From departure to arrival, an aircraft pushes back from the gate, taxis out to the runway, takes off, passes though many en route sectors in the air, lands, and finally taxis to the gate. At the gate, the aircraft waits for turn-around, after which it continues on to the next leg. The Fight Delay database was constructed to record movement information of every 2-consecutive-leg components for each aircraft.

The Flight Delay database was constructed using Microsoft SQL Server. It combines data from the Aviation System Performance Metrics (ASPM), National Convective Weather Detection (NCWD) databases, and BTS on-time performance database. In the ASPM database, each record contains information on one segment of flight, i.e. from the time (scheduled, planned and actual) the airframe pushes back from the gate to the time (scheduled, planned and actual) it enters the gate, depicted by the double-headed thick arrow in Figure **3.1**.

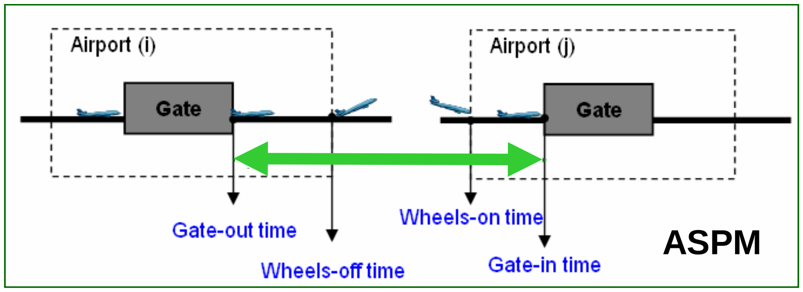


Figure **3.1**: ASPM Record Scope

The shortcoming of the way that the ASPM records the data for individual flights is the missing information from the previous leg. In this dissertation, a leg is defined as the process that an aircraft takes in traveling from one airport to another airport. The end points of a flight leg are gate in time at the origin airport and at the destination airport. If an aircraft traverses multiple airports in a day, it has a sequence of flight legs over a day.

The constructed Flight Delay database is connected by aircraft tail number. Therefore, its records contains information about two legs of flight as shown in Figure **3.2**, i.e., from the time (scheduled, planned and actual) the aircraft enters the gate at the previous leg to the time (scheduled, planned and actual) it enters the gate at the current leg.

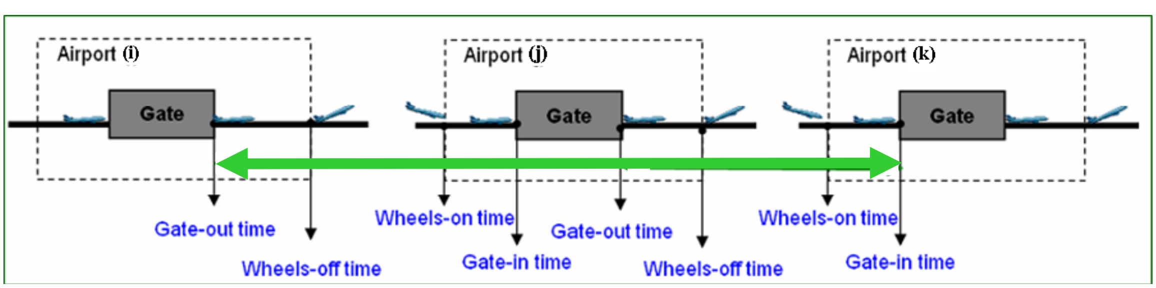


Figure **3.2**: Construction of Connected Database

Figure **3.3** shows a vehicle delay model that was developed to depict the occurrences of delays in a flight's itinerary. In Figure **3.3**, ACT represents actual, SCH represents scheduled, SEC represents second, IN means gate in, ON means wheels on, OFF means wheels off and OUT means gate out. The zero following SEC indicates the time of previous leg. This model divides delays in an itinerary of a flight leg into pieces according to flight segments. These delays were analyzed in order to gain understanding of the causal factors a flight may encounter in each segment.

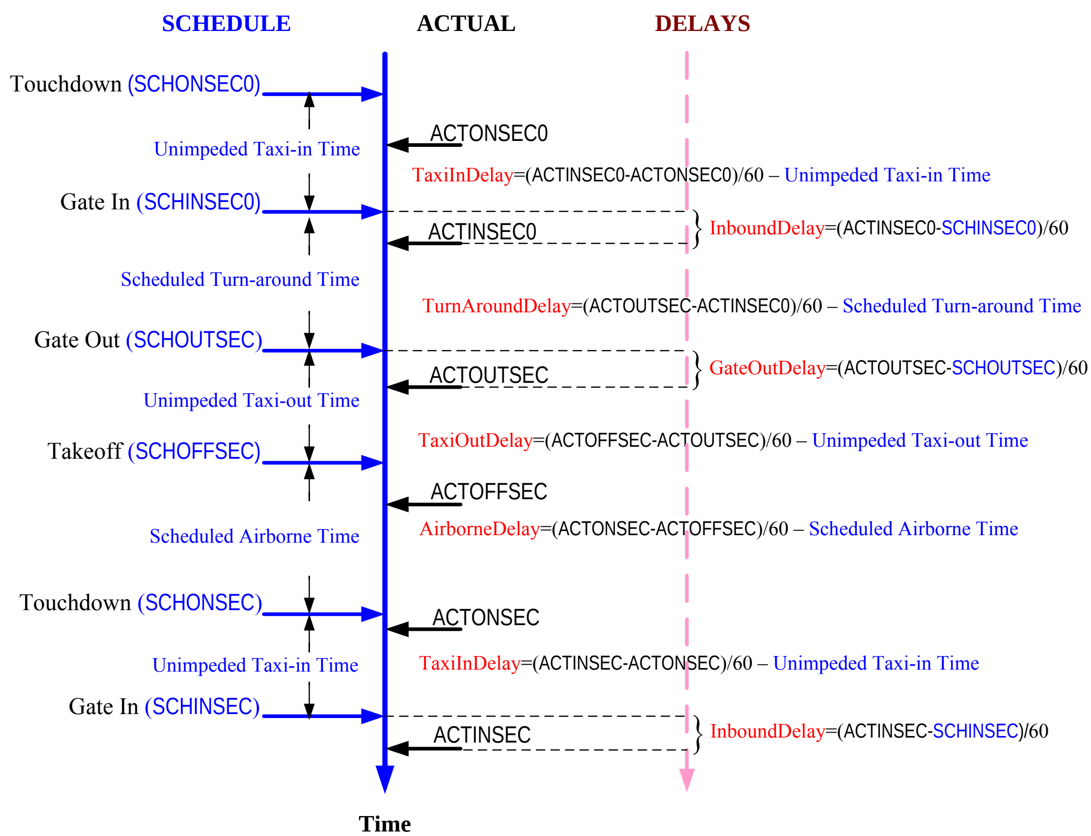


Figure **3.3**: Description of Vehicle Delay Model

In Figure **3.3**, the scheduled starting time of each flight segment is listed along the vertical time line in the section to the left. The discrepancy between two scheduled starting times is the scheduled segment time. The actual starting times of corresponding segment is on the right side of the time line. Between the adjacent actual starting times, the equation to calculate segment delay is shown. Segment delay is the difference between the actual segment time length and the scheduled time length. The dashed vertical line in the section to the right separates the segment delay and point delay. Inbound Delay and Gate-out delay are referred to as point delay because they are the difference of actual starting point of a segment and the corresponding scheduled starting point.

In a database, the primary key is a minimal set of attributes (columns) whose values uniquely identify an entity (record) in the set (Ramakrishnan and Gehrke 2000). The primary key of the Flight Delay database is the aircraft tail number, departure year and month, departure day and scheduled departure time in both hh:mm format and accumulated second since January 1st 1980.

Flight operation involves a great deal of uncertainty. Flight delay analysis is conducted at the lowest level of ATS. Existing research has shown that the flight delay has a different relationship to the causal factors at different O/D pairs. (Xu et al 2007). Having more than 1000 O/D pairs among 35 OEP airports, there is a need of more general models to analyze the delays in ATS. A higher level, meso-level, airport delay database is needed to gain insight into system delay.

3.1.3 Airport Delay Database

A higher level Airport Delay database was aggregated from the flight database. The primary key of Airport Delay database is airport ID, departure year and month, departure day and scheduled departure time (15-minute epoch).

Figure **3.4** and Figure **3.5** plot the aggregated delays of outbound flights from PHL and LGA in June, July and August 2005. These flights are recorded in BTS. Their previous leg information can be traced back in the database if the current leg is not the first leg of the day. These flights were scheduled to depart at the same 15-minute epoch are aggregated together for each epoch of day. The average Inbound Delay, Generated Delay, Early Arrival Gap, Absorbed Delay, and Wheels-off Delay throughout the day at Philadelphia International Airport (PHL) and LaGuardia Airport (LGA) were represented as stack bars. Each bar is a component of the average wheels-off delay (dot) for each flight scheduled to depart in a given 15 minute period. The average delay experienced during this period is the inbound delay (grey) plus the Generated Delay (black) plus the Early Arrival Gap (white) minus the Absorbed Delay (strip).

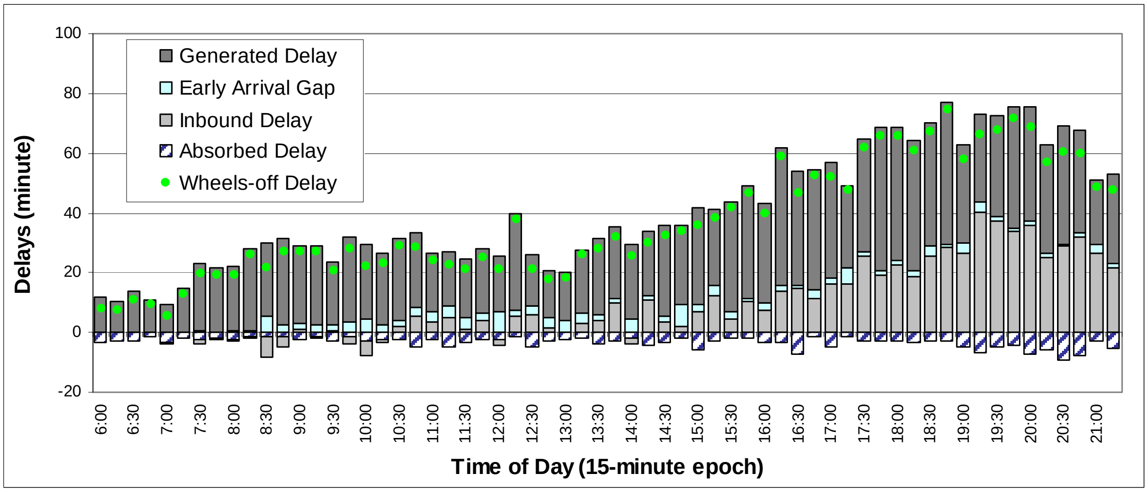


Figure **3.4**: Mean value of the components of the delays experienced by flights scheduled to depart in each 15 minute period for 3 summer months at PHL. The wheels-off Delay (dot) is the Generated Delay (black bar) plus the Early Arrival Gap (white bar) plus the Inbound Delay (grey bar) minus the Absorbed Delay (strip bar).

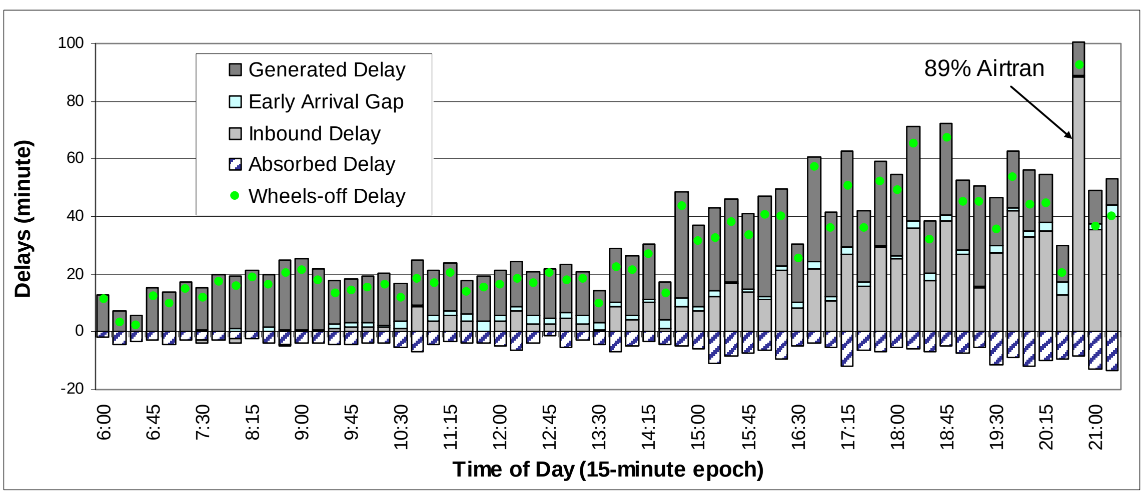


Figure **3.5**: Mean value of the components of the delays experienced by flights scheduled to depart in each 15 minute period for 3 summer months at LGA. 89% flight at the epoch 83 is from Airtran Airline.

PHL and LGA are both airports with high Airport Generated Delay. However, the magnitude of delays at LGA in Figure **3.5** fluctuates more than the magnitude of delays at PHL. The aggregated airport data reveals more general information about the airport performance. The Figures for 34 OEP airports are listed in Appendix C, which aggregates the average delay in 1-hour increments through out a day for summer 2006.

Airport Operation Center (AOC) personnel and Traffic Flow Management (TFM) Specialists have suggested that a model for predicting airport delay in 15 minute epochs would be useful. Hence, in the Airport Delay database, a day is divided into 96 15-minute epochs that are numbered from 0 to 95. The delays in each epoch are average value per flight for flights scheduled to depart in that epoch. For example, the Airport Generated Delay of $epoch_i$ is the average of Generated Delays of all flights scheduled to depart at $epoch_i$; the Inbound Delay of $epoch_i$ is the average of Inbound Delays of same flights scheduled to depart at $epoch_i$. Along with the data of potential factors causing delay is the Airport Delay database.

The Airport Delay database contains only the connecting flights in the BTS database. A flight is coded as a connecting flight if its previous leg information is not absent and its tail number is in the correct format. If the flight is the first flight of a day then it is also a connecting flight but with zero Inbound Delay.

The data associated with extreme value of delays in the Airport Delay database was excluded from the data set used to develop and validate the airport delay model. Examples of excluded values are Inbound Delays or Airport Generated Delays exceeding 270 minutes (4.5 hours); Absorbed Delays longer than 3 hours; previous leg flight arrived more than 45 minutes early. Each extreme scenario accounts for less than 0.1% of total data.

3.2 Factor Collection

Janic identified the contributions of two major causes for flight delays in U.S: bad weather and congestion (Janic 2005). Bad weather alone accounts for 70-75% of flight delays. Airport and airspace congestion accounts for another 20-30%. However, congestion problems are often associated with bad weather. Some researchers have tried to estimate flight delays from these two types of factors and their interaction.

Lamon (2001) estimated the number of Operations Network (OPSNET) delays at the 55 busiest commercial airports using 2 factors, percentages of convective weather coverage and total number of arrivals and departures. OPSNET delay refers to the delays of 15 minutes or more experienced by Instrument Flight Rule (IFR) flights provided through FAA (FAA 1). Callaham et al. (2001) derived a composite NAS performance measure as a response variable and regressed on a categorical variable with respect to schedule traffic level and pattern and another categorical variable with respect to severe weather location and extent. Although, the response variables describe aggregated NAS performance from different perspectives in the research of Lamon and Callaham, respectively, the predictor variables in their models are all about weather related variables, congestion related variables and their interactions.

Similar but not limited to the factors in the research of Lamon and Callaham, the factors collected in this dissertation research involve 5 groups: weather related factors, traffic related factors, airline related factors, traffic flow management related factors, and other factors which cannot be categorized into the first 4 groups. These factors were defined based on the literature related to aviation delays and either selected or derived from the available databases.

3.2.1 Weather Related Factors

Airport Weather

Weather is a major contributor to large delays. Airport Arrival Rate (AAR) (it is also referred to Airport Acceptance Rate) and Airport Departure Rate (ADR) depend mostly on weather (and on demand as well) (Hansen and Bolic 2001). Airlines usually schedule for fair weather, so any decrease in AAR/ADR caused by weather can lead to delays. Factors related to weather in the existing literature are defined as convective weather, reduced ceiling and visibility, wind and corresponding AAR /ADR at local and destination airport respectively (e.g. Allan et al 2001, Evans and Clark 2005, Hansen and Hsiao 2005 etc.). The Quarter Hour Airport Report from FAA database, ASPM, records weather condition for each 15-minute epoch. In addition to ASPM, the categorical weather data for the local airport from the National Convective Weather Detection (NCWD) is also collected. The states of local weather in NCWD include thunderstorm, heavy-rain, rain, high-wind, wind, low-ceiling, low-visibility and none. Table **3.1** gives the ordinal representation for each state of local weather.

Table **3.1**: Ordinal Representation of Local Weather

States of Local Weather in NEWD	Ordinal number
NONE	0
WIND	1
RAIN	2
LOW_VIS	3
LOW_CLG	4
LOW_MDT_WIND	5
H_RAIN	6
H_WIND	7
TSTORM	8

The weather related variables in ASPM are as follows:

- WindSpeed: Wind speed (Knots) is recorded every 15 minutes for an airport.
- WindDirection: Wind direction (angle) is recorded every 15 minutes for an airport.
- Ceiling: The heights in 100 feet above the Earth's surface of the lowest level of clouds.
- Visibility. It is the ability, as determined by atmospheric condition and expressed in units of distance, to see and identify prominent unlighted objects by day and prominent lighted objects by night, which is measured in Statute Miles for Flight Operations (Pilot/Controller glossary).
- Weather: categorical variable with states, thunderstorm, heavy rain, rain, high wind, wind, low ceiling, low visibility and none, after different time period, from NCWD.

- IFR/VFR: airport visual condition. IFR refers to Instrument Meteorological Conditions. VFR refers to Visual Meteorological Conditions.
- Runway: Runway configuration (Arr | Dep) represents the arrival and departure runways in use at the time of the event. Different runway combination requires different AAR and has different influence on delay. In this research, the normalization was applied on the Runway Configuration. The numerical value 1, 2 and 3 were assigned to groups of Runway Configuration based on the mean taxi-out delay of that group. The number 1 is assigned to the group of Runway Configuration which has the lowest mean taxi-out delay.
- AAR: refers to the Airport Arrival Rate in the ASPM (also see the Airport Acceptance Rate). The AAR represents the reported maximum number of arriving aircraft that a facility can handle per unit of time (i.e., quarter hour or hour), reported to the Operational Information System (OIS).
- ADR: Airport Departure Rate in ASPM represents the reported maximum number of departing aircraft that a facility can handle per unit of time (quarter hour).

En-route Severe Weather Report

Callaham et al. (2001) developed a Weather Impacted Traffic Index (WITI) to normalize for the potential impact of severe convective weather on en-route performance. The en-route WITI was calculated by overlaying a grid on the U.S. and assigning a weight to each cell based on the potential impact of air traffic performance from the convective weather that occurred in that cell. The computation of WITI was later extended by Chatterji and Sridhar (2005) to include the regions around severe weather cells. The extended WITI was used to explain the total daily delays in the NAS.

In this research, a more detailed variable is needed to represent the en-route weather condition since the response variable is delay in a 15-minute epoch. The NCWD records the total number of severe weather reports on each flow and airport hourly. The number of severe weather reports during a flight's scheduled or actual en-route time was calculated from the NCWD as a surrogate for this flight's en-route airspace weather condition.

Figure **3.6** illustrates the calculation of the total number of en-route severe weather reports for a flight which departed at $h_i : x$ and arrived at $h_j : y$, where h represents hour, x and y represent minutes. The calculation includes three parts:

(1) The portion of the number of severe weather reports at the first hour h_i, which is $\left(1-\frac{x}{60}\right)$*Report$_i$. Report$_i$ represents the number of severe weather reports in hour h_i.

(2) The total number of reports in the time interval between h_{i+1} and h_j.

(3) The portion of reports at the last hour h_j, which is $\frac{y}{60}$*Report$_j$.

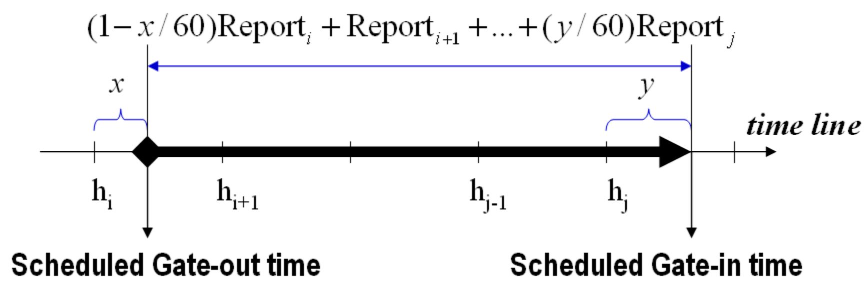


Figure **3.6**: Calculation of En-route Severe Weather Report during Scheduled En-route Time

The Zulu time in the NCWD was converted into airport local time in order to match with the flight in ASPM (*http://www.grc.nasa.gov/WWW/MAEL/ag/wtz1.gif*).

3.2.2 Traffic Related Factors

Ratio of Operation Demand and Airport Capacity at Scheduled Departure Time (ρ)

Researchers formulated several kinds of variables to represent the operation demand, airport capacity, and the relationship between demand and capacity, such as departure demand and queue size (Idris et al 2002), levels of traffic (Hoffman 2007), arrival demand and total flight operations (Hansen and Zhang 2004), arrival queue and volume (Hansen and Hsiao 2005), airport throughput (Welch and Ahmed 2003), and rho (ρ), the ratio of number of scheduled operations to capacity (Ball et al 2006). Each of these tries to quantitatively describe the traffic at an airport from a different perspective.

In this research, airport operation demand is divided into departure demand and arrival demand, and airport capacity is divided into departure throughput and arrival throughput. In a specific time window, departure demand counts the number of flights scheduled or actually pushed back from gate; arrival demand counts the number of flights scheduled to enter the gate; throughput counts the number of actual wheels-offs for departure and number of actual gate-ins for arrival.

For an individual flight *i*, the ratio of departure demand and departure throughput, abbreviated as Departure Demand Ratio, is defined as the scheduled (or actual) departure demand divided by departure throughput in a specific time window of the scheduled (or actual) departure time of this flight at its origin airport or destination airport (as in Equation **3.1**). These ratios were calculated in time windows 15-minute and 30-minute around scheduled and actual operation time for departure and arrival separately. Note: the Departure Demand Ratio_ADR$_i$ denotes the ratio of departure demand and the declared Airport Departure Rate (ADR) by FAA.

$$\text{Departure Demand Ratio}_i = \frac{\text{Departure Demand}_i}{\text{Departure Throughput}_i} \qquad \textbf{3.1}$$

$$\text{Departure Demand Ratio\_ADR}_i = \frac{\text{Departure Demand}_i}{\text{ADR}_i}$$

The departure demand includes the number of scheduled departures (calculated from ASPM) and cancelled departures (calculated from BTS) (Equation **3.2)**.

$$\text{Departure Demand}_i = \text{Scheduled Departures}_i + \text{Cancelled Departures}_i \qquad \textbf{3.2}$$

The schedule departures for flight *i* in a 30-minute time window of its Scheduled Departure Time t_i is the summation of flights whose scheduled departure time is within the time interval of ($t_i - 15$ minutes, $t_i + 15$ minutes). This number is calculated from the

ASPM. The records in the ASPM are flights actually flown in the NAS; the cancelled flights are not included. The number of cancelled flights whose scheduled departure time is within the time interval of ($t_i - 15$ minutes, $t_i + 15$ minutes) is collected from the BTS database. However, the BTS database only contains flights reported by certified U.S. air carriers that account for at least one percent of domestic scheduled passenger revenues (BTS sources). There are 20 carriers in the BTS database in the 3-month period 2005, while there are records for 534 carriers it the ASPM. Therefore, the calculated cancellation from the BTS is just a subset of actual cancellations.

The departure capacity is calculated as the airport departure throughput, which is the number of flights whose actual wheels-off time is within the time interval ($t_i - 15$ minutes, $t_i + 15$ minutes).

For each 15-minute epoch, the ratio of departure demand and departure capacity, abbreviated to Departure Demand Ratio$_{30}$, is the average value of Departure Demand Ratio$_i$ in 30-minute window for flights whose scheduled departure time is in that epoch (Equation **3.3)**.

$$\text{Departure Demand Ratio} = \frac{1}{n}\sum_{i=1}^{n} \text{Departure Demand Ratio}_i \qquad \textbf{3.3}$$

Arrival Demand Ratio is defined as the scheduled arrival demand divided by arrival throughput in Equation **3.4**. Arrival Demand Ratio$_{30}$ denotes the average Arrival Demand Ratio in 30-minute window.

$$\text{Arrival Demand}_i = \text{Scheduled Arrivals}_i + \text{Cancelled Arrivals}_i$$

$$\text{Arrival Demand Ratio}_i = \frac{\text{Arrival Demand}_i}{\text{Arrival Thrroughput}_i} \qquad \textbf{3.4}$$

$$\text{Arrival Demand Ratio} = \frac{1}{n}\sum_{i=1}^{n} \text{Arrival Demand Ratio}_i$$

Arrival Demand Ratios and Departure Demand Ratios at the origin and destination airport are all analyzed in this research.

In the situation where the departure throughput is zero, Departure Demand Ratio is set equal to the departure demand.

The variable denoted as Departure Demand Ratio_ADR30(or 15) represents the ratio of departure demand and Airport Departure Rate (ADR) in 30 (or 15) minute time window. Similarly, the variable denoted as Arrival Demand Ratio_AAR30(or 15) represents the ratio of arrival demand and Airport Arrival Rate (AAR) in 30 (or 15) minute time window. The variable denoted as Departure Ratio_15 represents the ratio of departure demand and the maximum value of ADR and departure throughput in a 15-minute epoch.

Scheduled Departure Time

Hsiao and Hansen (2006) estimated the relationship between time of day and daily NAS delay. Beatty et. al (1999) also identified the different impact of initial delay at different time of day. In this paper, Scheduled Departure Time is defined in 15-minute epoch from 0 to 95 for a day. We only keep the data where the epoch is between 24 and 87 (6:00AM to 9:59PM).

Inbound delay

The Inbound Delay of a flight is the delay accumulated from upstream airports and en-route legs, i.e., the propagated delay from previous legs. The value of Inbound Delay is the difference between actual gate-in time and the scheduled gate-in time. The Inbound Delay of a 15-minute epoch is calculated by average Inbound Delays of all flights scheduled to depart to other 33 OEP airports at that epoch.

3.2.3 Airline Related Factors

Different airlines have different operations and scheduling strategies (Beatty et al. 1999), and the same airline has different strategies at hub and spoke airports. In the Flight Delay database, different airlines are identified by different airline code; while in the Airport Delay database, there is no distinction between airlines since individual airlines' data was merged together into an aggregate value for each 15-minute epoch.

Scheduled Turn-around Time

Wang et al defined turn around time as the time between an aircraft's arrival and subsequent departure at the same airport (Wang et al 2003). They found that ample slack and flight time allowance in turn-around time and flight time can absorb most delays for subsequent flights. Vigneau referred to Scheduled Turn-around Time as station stop time in his paper and found it is an important variable for flight departure delay (Vigneau 2003).

Scheduled turn-around time in the CATSR Delay database is defined as the length of time from scheduled gate-in from the previous leg to the scheduled gate-out time of the current leg. At an epoch, if all flights' information for the previous leg is missing and the current leg is not the initial leg of the day, the average Scheduled Turn-around Time of that airport is used.

Leg Number

Schaefer and Miller (2001) found that the propagated delay is significant for the 1st leg after leaving an Instrument Meterological Condition (IMC) airport and then damps out. We define a variable, Leg Number, to represent the position of a flight in its whole day itinerary. In the Flight Delay database, each aircraft's itinerary for a day can be tracked by its tail number and its scheduled departure time. The Leg Number starts at one.

Beatty et al (1999) traced individual aircraft's itinerary throughout a day using the data from American Airline, then calculated Delay Multipliers for initial delay that occurred at different time of day. Delay Multiplier represents the delay impact on the operation schedule as a whole. They found that later initial delay corresponds to smaller values of Delay Multiplier. Later time of day is usually associated with larger Leg Number especially for airlines having tightly connected schedules such as American Airlines. Hence, Leg Number is a potential factor for delays.

Carrier Delay

Carrier Delay records the length of delay (in minute) caused by aircraft cleaning, aircraft damage, awaiting the arrival of connecting passengers or crew, baggage, bird strike, cargo loading, catering, computer, outage--carrier equipment, crew legality (pilot or attendant rest), damage by hazardous goods, engineering inspection, fueling, handling disabled passengers, late crew, lavatory servicing, maintenance, oversales, potable water servicing, removal of unruly passengers, slow boarding or seating, stowing carry-on baggage, weight and balance delays (DOT Report 2000). The causes of these delays are the circumstances considered within the airline's control by FAA.

For each 15-minute epoch, the Carrier Delay is the average value of Carrier Delay of individual flights scheduled to push back from gate at that time interval.

Aircraft Substitution (or Swapping Aircraft Rate)

Airlines substitute different aircraft to minimize the impact of long arrival delays (Beatty et al 1999). The criterion to identify a Swapped (substituted) airframe is set by comparing the airframe's scheduled gate-in time from the previous leg to the scheduled gate-out time at the current leg.

For an individual flight *i*, Swapped Aircraft is an indicator variable. Swapped Aircraft is equal to one when its scheduled departure time at current leg is earlier than its scheduled arrival time from the previous leg; otherwise Swapped Aircraft is zero.

For an epoch at an airport in Airport Delay database, Swap Aircraft Rate is a numerical variable representing the proportion of aircrafts which were not initially scheduled to all scheduled flights in the Flight Delay database.

Cancellation (arrival and departure)

A cancelled flight is "a flight that was listed in a carrier's computer reservation system during the seven calendar days prior to scheduled departure but was not operated" (BTS Glossary). In the Airport Delay database, the cancelled departures (arrivals) are calculated as the count of cancelled flights in the BTS database in 15-minute epochs.

Load Factor

In Vigneau's research of flight delays in Europe (2003), the passenger load of an aircraft is an important factor for departure delay. Since there is no data available for passenger load, the weight class and number of seats of an aircraft are collected to reflect the complexity of check-in, boarding and unboarding operations.

Weight class and number of seat were obtained from the ETMSC database. The matching of ETMSC with ASPM is done by linking the aircraft types in two databases.

3.2.4 Traffic Flow Management Related Factors

GDP Holding Time

The Ground Delay Program (GDP) is issued to reduce the volume of inbound flights at a destination airport when the scheduled demand exceeds the capacity of that airport. Over-scheduling and severe weather at destination airport can both result in a GDP at an origin airport (Allan et al 2001). The GDP Holding Time of an individual flight measures its assigned holding time on the ground of origin airport by Air Traffic Control before it departs for its destination. Idris et al (2002) pointed out that the downstream restrictions including the GDP affect an aircraft's taxi out time. Hansen and Zhang (2004) estimated impact of GDP Holding Time (called EDCT holding in their paper) on the arrival delay at LGA.

In this research, the GDP Holding Time is calculated from the ASPM Individual Flight database. For a single flight, the GDP Holding Time is the time elapse between the scheduled wheels-off time ($SchOffSec$) and Estimated Departure Clearance Time (EDCT) wheels-off time ($EDCTOffSec$).

The first formula for GDP Holding Time is defined as Equation **3.5** :

$$GDPtime_i 1 = \begin{cases} (EDCTOffSec - SchOffSec)/60 & EDCTOffSec <= ActOffSec \\ (ActOffSec - SchOffSec)/60 & otherwise \end{cases} \qquad \mathbf{3.5}$$

Some adjustment was made for the situation where the aircraft's actual take-off time ($ActOffSec$ or $ActOffTm$) was earlier than the assigned GDP wheels off time. The actual take-off time is treated as the ending point of GDP Holding Time in the CATSR database. In the ASPM database, *EDCTOffSec* and *EDCTOffTm* both record the assigned take off time by GDP. *EDCTOffSec* records the GMT seconds of EDCT wheels off time since 1/1/1980; *EDCTOffTm* records EDCT wheels off time in the format HH:MM. These two records do not always agree with each other.

The second formula for GDP Holding Time is defined as Equation **3.6** :

$$GDPtime_i 2 = \begin{cases} left(EDCTOffTm,2)*60 + right(EDCTOffTm,2) \\ \qquad - left(SchOffTm,2)*60 - right(SchOffTm,2) \\ \\ left(ActOffTm,2)*60 + right(ActOffTm,2) \\ \qquad - left(SchOffTm,2)*60 - right(SchOffTm,2) \end{cases} \tag{3.6}$$

$$\text{Where, } left(ActOffTm,2)*60 + right(ActOffTm,2) < left(EDCTOffTm,2)*60 + right(EDCTOffTm,2)$$

If GDP assigned wheels off time is on the next day, the adjustment of secondary formula is in Equation **3.7** :

$$GDPtime_i 2 = [left(EDCTOffTm,2)+24]*60 + right(EDCTOffTm,2) - left(SchOffTm,2)*60 - right(SchOffTm,2) \tag{3.7}$$

Given two sets of GDP Holding Time, $GDPtime1$ and $GDPtime2$, both calculated from the same database, the value closer to the Airport Generated Delay (*gendla*) is selected as the GDP Holding Time. This reflects the assumption that the Generated Delay is a direct result of GDP Holding Time (Equation **3.8**).

$$GDPtime_i = \begin{cases} GDPtime_i 1 & \text{when } \left| GDPtime_i 1 - gendla_i \right| <= \left| GDPtime_i 2 - gendla_i \right| \\ GDPtime_i 2 & \text{when} \left| GDPtime_i 1 - gendla_i \right| >= \left| GDPtime_i 2 - gendla_i \right| \end{cases} \quad \mathbf{3.8}$$

In this research, the factor used in the model to estimate the average delay in each 15-minute epoch is the average GDP Holding Time of all flight scheduled to push back from gate at that time interval.

3.2.5 Other Factors

Security Delay

Security Delay is defined as “delays caused by evacuation of a terminal or concourse, re-boarding of aircraft because of security breach, inoperative screening equipment and/or long lines in excess of 29 minutes at screening areas”. (BTS 3)

Out-bound Leg Distance

Non-stop distance has been found as an important categorical factor in flight arrival delay in the research of Abdel-Aty et al. (2007). It was included in this research to investigate if it has any impact on airport operation.

CHAPTER 4

MULTI-FACTOR MODELS FOR AIRPORT DELAY

The delay factors have been described explicitly in Chapter 3. These factors are referred to as causal factors in the literature by the experts in the field. Theoretically, causality or causation is

> "a process linking two or more events or states of affairs so that one brings about or produces the other. One event is the cause of another if (a) the event occurs prior to the effect, (b) there is an invariant conjunction of the two events and (c) there is an underlying mechanism or physical structure attesting to the necessity of the conjunction." (Web Dictionary of Cybernetics and Systems, available at http://pespmc1.vub.ac.be/asc/indexASC.html).

To establish a firm empirical argument for a causal relationship, (b) and (c) have to be justified by empirical data from controlled experiments. Due to the lack of empirical data (mainly because it is practically infeasible to do the necessary experiments), the causal factors studied in the existing literature are defined based on expert judgments.

In a complex interrelated system, NAS, the causes of delay are multiple and rarely deterministic. The objective of this dissertation research is not to demonstrate causality, but to distinguish the most important factors from other researchers' study and to predict delays from the selected factors.

This chapter describes a method to identify the important factors and predict airport delay from these factors deriving from regression analysis.

Section **2.3.2.2.1** discussed two sources of problems with regression model: data problems and model problems (Freund et al 2006). The data process described in chapter 3 has tried to collect all possible factors and not miss important predictors related to airport delay. Reducing influential observations is also achieved by removing the extreme long delays from Flight Delay database.

The method described in this chapter is designed to overcome the model's problems in regression analysis.

One kind of problematic model is an inadequate model. Inadequate models miss important predictors and/or incorrectly specify the relationships. An example of this model would be representing a non-linear relationship as linear. The first part of this chapter describes how and why the piece-wise linear regression model was selected from various kinds of regression methods.

Another kind of problematic model is one which is too complex or overspecified, which includes too many explanatory variables and/or too many higher-order terms. The existence of collinearity among predictor variables causes the estimated regression coefficients to have inflated standard errors and also makes it difficult to distinguish their individual influence on the response variable. The second part of this chapter describes the approaches used to deal with collinearity and overfitting.

4.1 Deriving a Multi-Factor Model

National Airspace System (NAS) is a complex adaptive system, and complex adaptive systems are highly non-linear (Donohue 2003). Non-linear relationships can be modeled by various types of regression methods. The selected regression model should be able to represent the complex non-linear relationship in a compact format.

4.1.1 Overview of Regression Methods

This study compared OLS (discussed in Chapter 2) with several other regression methods. The other methods considered in this study include two local regression methods (Kernel regression and Nearest-neighbor regression), projection pursuit regression, and three adaptive nonlinear methods (CART, MARS, and MART). These methods are discussed below.

Local regression methods: Kernel regression and Nearest-neighbor

The basic idea of local regression is that one does not try to find a single model to fit all of the observations in a data set (and fit the model using all of the data). Instead to predict a response value for a given set of predictors, only observations having similar predictor values are used. That is, to predict the response for a vector of predictor variables, **x**, only observations having predictor values in a neighborhood of **x** are used. When the space of predictors has high dimension, this method can create a large bias in the predictions (Hastie et al. 2001).

Projection Pursuit Regression (PPR)

PPR is a type of additive model. It is additive not in the original feature space, but in a space of derived features. The format of the model is

$$f(\mathbf{x}) = \sum_{m=1}^{M} g_m\left(\varpi_m{}^T\mathbf{x}\right), \qquad \mathbf{4.1}$$

where the derived features, the $V_m = \varpi_m{}^T\mathbf{x}$ $(m = 1, \cdots, M)$, are projections of **x** onto various directions. An iterative method is used to search for values of the ϖ_m and the functions g_m $(m = 1, \cdots, M)$ which provide a good fit. For this reason, this method is called Projection Pursuit. The parameter M can be arbitrarily large, and any continuous function in the sample space can be approximated arbitrarily well given adequate data. A well fitted regression model can be constructed in this way. However, interpreting the fitted model and gaining an understanding of the phenomenon being modeled can be extremely difficult (Hastie et al. 2001).

CART

Classification And Regression Trees (CART) is a decision-tree based regression method (Breiman et al. 1984). CART can be used for both classification and regression.

In this research, CART is used to grow regression trees. The measurement space Ξ is recursively portioned with a sequence of splits on the predictor variables. The partitioned space can be represented with a tree-like structure. For each of the nodes (corresponding to the disjoint regions of the partition of the space) of the tree, the estimate of the response variable is a value, $\widehat{g(\mathbf{x})}$. The same estimate is used for all $\mathbf{x}$ within the range of a node. The splits are selected in order to minimize $\sum_{i=1}^{N}\left[y_i - g(\mathbf{x}_i)\right]^2$ or $\sum_{i=1}^{N}\left|y_i - g(\mathbf{x}_i)\right|$. A split is chosen if it produces the greatest possible reduction of $\sum_{i=1}^{N}\left[y_i - g(\mathbf{x}_i)\right]^2$ or $\sum_{i=1}^{N}\left|y_i - g(\mathbf{x}_i)\right|$. The final tree is selected based on the prediction accuracy achieved by a test sample or by cross-validation estimates.

MARS

Multivariate Adaptive Regression Splines (MARS) fits a series of continuous piece-wise linear functions called splines. A spline is a piecewise polynomial function that passes through every data point (interpolating splines) or close to the data point (smoothing splines). A linear spline is the simplest type of spline. The MARS model combines splines additively.

Starting with a constant in the model, MARS adds one variable-knot combination at a time. Graphically, the added combination is a line bent at a certain value t (called a knot). Two straight pieces of this bent line have the form $(x-t)_+ = \begin{cases} x-t, & \text{if } x > t, \\ 0, & \text{otherwise,} \end{cases}$ and $(t-x)_+ = \begin{cases} t-x, & \text{if } x < t, \\ 0, & \text{otherwise.} \end{cases}$ These functions are called basis functions. At each step in the model building phase, a basis function pair is added by choosing a predictor variable to assume the role of x and selecting a value for the knot, t. The selected basis function is the one that gives the greatest amount of improvement, where improvement is defined as the greatest reduction in the sum of squared errors. Terms are added into the model until it grows so complex that it overfits the data. Then MARS removes one basis function at a time to create a sequence of pruned models until all basis functions except the first constant are removed. Each time, the removed function is the one that contributes the smallest increase in the sum of squared residuals. The final model is selected through cross-validation from the sequence of created models.

MART

Multiple Additive Regression Trees (MART) model is a collection of weighted and summed trees. MART is similar in spirit to a long series expansion (such as a Fourier or Taylor's series) – a sum of factors that becomes progressively more accurate as the expansion continues. The expansion can be written as:

$$F(\boldsymbol{X}) = F_0 + \beta_1 T_1(\boldsymbol{X}) + \beta_2 T_2(\boldsymbol{X}) + \cdots + \beta_M T_M(\boldsymbol{X}). \qquad \textbf{4.2}$$

The terminal nodes of each tree in the sequence are selected by a greedy, top-down recursive partitioning algorithm. Each terminal node corresponds to a disjoint region of the partition of the space. A slow-learning procedure is used, for which at each step the coefficient for the new tree added is decreased in magnitude from what it should be if the model-building process were to stop at that step. This creates a set of residuals looking less random than they would be if a larger coefficient was used, so that the next tree added can act to capture the pattern in the residuals. When a large number of such trees are superimposed, the response surface can be well approximated.

4.1.2 Regression Methods Comparison

A comparison of regression methods was done to select an appropriate regression method which can well approximate the non-linear relationships and create models which are accurate preditors. Also, it is desired to find a method for which the model building process requires less effort and is more automatic, since 68 models are needed for 34 airports.

Sample Data:

Airport Delay involves two flight segments, turn-around and taxi-out. Xu et al. used Bayesian network models to infer that Taxi-out Delay is one of the major contributors to arrival delay at destination (2007). Taxi-out delay has been estimated by non-linear models (Irids et al. 2002). A regression model which can accurately estimate the non-linear relationships of Taxi-out Delay and its predictors should be a reasonable choice to predict Airport Delay.

Furthermore, Taxi-out Delay data can be collected more easily than Turn-around Delay since it involves only one leg information of flights. The taxi out delay data from the ASPM Quarter Hour Report for ORD from December 1st 2003 to February 29th 2004 was collected to compare the prediction power of the regression methods. Observations from early morning or with missing variables were removed from analysis. The causal factors were selected from the existing literature.

The data sample was divided into 3 parts: training data, testing data and generalization data. The testing data and generalization data were randomly selected from complete data set. Each of them accounts for 20% of the data, with the remaining 60% serving as the training data.

Comparison Design:

The regression models compared include Ordinary Least Square Regression (OLS), local regression methods Nearest Neighbor Regression and Kernel Regression, and computer-intensive regression methods (PPR, CART, MARS and MART). Defining Mean Squared Prediction Error (MSPE) as $MSPE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2$,

where, y_i : the actual delay in testing sample,

$\hat{y}_i$: the model estimated delay using data from testing sample,

n : the number of cases in the testing sample.

The MSPE is used as the criterion to compare the model prediction performance.

The comparison was performed in two phases.

- Phase I: regression models were built from the training data. Then, the testing data was used to assess the estimation accuracy of the models and select the best settings and the best model through MSPE.

  - For the more traditional regression method, the testing data was used to choose between OLS regression with stepwise variable selection and a shrinkage method (Lasso).
  - For local regressions, the testing data was used to choose the number of nearest neighbors and the shape of the kernel.
  - For PPR, the testing data was used to select number of directions (M).
  - For CART, the testing data was used to decide whether or not linear combinations of variables should be used.
  - For MARS, the testing data was used to choose maximum number of basis functions and the maximum number of interaction terms allowed in the model.
  - For MART, the testing data was used to determine the number of nodes per tree.

- Phase II, the training data and testing data of Phase I were combined together to form a larger size training set (learning data) and then update the parameters of model learned during Phase I. Generalization data set is applied on Phase II models in order to get unbiased estimates of the generalization errors of the models selected.

Comparison Results

OLS regression is relatively good provided that the error term distribution is normal or "well behaved" (i.e., not heavy-tailed and not highly skewed). If the relationship between a predictor and the response is sufficiently nonlinear, transforming the predictor or adding higher-order terms involving the predictor, may appreciably improve performance. However, this can be very complicated when there are a lot of explanatory variables. In fact, the variable selection and transformation were very time consuming.

Table **4.1** reports estimates of the MSPE for the different methods. A smaller MSPE means a better prediction. The estimated standard errors (SE) of MSPE is calculated as $\sqrt{\frac{\sum (w_i - \bar{w})^2}{m-1}} \Big/ \sqrt{m}$, where w_i is squared prediction error for the *i*th case, $\bar{w}$ is the sample mean of squared prediction errors, and m is number of cases.

Table **4.1**: MSPE of Learned Regression Model

Model:	Phase I MSPE	Rank	Phase II MSPE	Rank
OLS	0.0161	3	0.0189	7
nearest neighbor	0.0300	6	0.0176	5
kernel	0.0623	7	0.0174	4
PPR	0.0169	4	0.0170	2
CART	0.0189	5	0.0184	6
MARS	**0.0154**	**2**	**0.0173**	**3**
MART	0.0124	1	0.0120	1

The results from MART are the best in both phases. MARS is the second best in the phase I, and the third best in phase II. However, since the estimated standard error of the MSPE of MARS is 0.0012, the observed difference in performance between MARS and PPR is negligible. OLS is the third best in phase I but the worst in phase II.

A common problem for MART and PPR is model interpretation. Overall, the MARS model is the appropriate choice because: (1) the model building process is automatic and fast, (2) it can provide accurate predictions for non-linear phenomena, and (3) the developed model is relatively easy to use and interpret.

4.2 Factor Selection

Based on the comparison results in the Section 4.1, regression models created using Multivariate Adaptive Regression Splines (MARS) were selected to estimate the non-linear relationship between airport delays and their influencing factors.

MARS is a commercial software developed by Salford Systems. It is an adaptive data-driven method. MARS systematically searches for the locations and number of knots and the interactions between variables. The nonlinear contributions of variables are approximated through a series of continuous piecewise linear splines learned by MARS.

The causal factors described in Chapter 2 were provided as input factors to MARS. The Airport Generated Delay and Airport Absorbed Delay are response variables. In situations where there are too many explanatory variables, the collinearity among explanatory variables and model over-specification are a major problem in regression analysis. For example, airport weather condition affects airport runway configuration and operation capacity. Weather condition, runway configuration and AAR/ADR are all related to airport delay. Including more predictor variables in the model may reduce prediction bias but increase prediction variance. Including fewer predictor variables may reduce prediction variance but increase bias. MARS's strategy of variable selection is to deliberately overfit a model first and then prune away parts which cause the least increases in the sum of squared residuals when they are removed. Cross-validation is used to select a final model from a generated sequence of different sized models.

At the initial part of this research, MARS was used to find models (Generated Delay model and Absorbed Delay model) for each of 34 airports. However, these models have different predictors and it is hard to analyze the factor's impact on delays collectively. In order to identify a more general relationship for airport delays across 34 OEP airports, the criteria of minimum sum of squared prediction errors can be slightly relaxed. If the hypothesis that prediction errors from two different sized models follow the same distribution cannot be rejected through statistical test, the models will be considered to have the same or nearly the same prediction power and the simple model will be favored.

4.2.1 Steps of Variable Selection

The historical data from each of the 34 OEP airports was separated into three sets: training data for creating the model, testing data for selecting the factors for the simple models, which do not seem to appreciably underfit the data, and validation data to estimate the prediction accuracy of the final model.

In the process of evaluating regression methods, the taxi-out data was randomly separated into three sets. For construction of models for airport delay, the data were separated into three sets by time. The data from June and July 2005 were treated as training data for predicting Airport Delays. The data from the first 15 days in August 2005 was treated as testing data. The last 15 days (excluding the 31st) in August 2005 was reserved for validation as described in Chapter 6. For each response variable – Generated or Absorbed Delay at one of 34 OEP airports – a set of predictor variables was selected and a model was generated according to the following steps:

i. Begin with all independent variables, using MARS to build a full-size regression model from the training data for each airport.
    a. First, apply a square root transformation to the response variable to symmetrize the residual distribution. The scatter plot and Quantile-Quantile plots of residuals for the untransformed response variable and transformed variable (square root transformation) are given in Appendix Figure **A.1**-**A.8**.

b. Next, obtain the residuals from the training data. $R_{full(i)} = y_{train(i)} - \hat{y}_{train(i)}$ where, $y_{train(i)}$ is the true delay value of case i in training sample, $\hat{y}_{train(i)}$ is the estimated delay value of case i calculated from the full-size model.

ii. Apply full-size model on testing data set to obtain prediction error $E_{full(i)} = y_{test(i)} - \hat{y}_{test(i)}$ where, $y_{test(i)}$ is the true delay value of case i in testing sample, and $\hat{y}_{test(i)}$ is the estimated delay value of case i calculated from the full-size model.

iii. Create a list of the selected variables from all 34 OEP airports. Count the number of times each variable appears among the 34 airports. The maximum appearance count is 34.

iv. Keep the variables whose appearance count is more than 17 (half of the total number of airports) as the factors to build a baseline reduced-size MARS model.

v. Calculate residuals $R_{reduce(i)}$ for the reduced-size model on training data for each airport.

vi. Apply reduced-size model on testing data and get prediction error $E_{reduce(i)}$.

vii. Conduct paired t tests to test the hypothesis that the mean of squared residuals from the regression and the squared prediction errors from full is greater or equal to the reduced size models. Conduct Kolmogorov-Smirnov tests to test the hypothesis that the squared values of residuals from the regression and the prediction errors from the full and reduced size models have the same distribution.

For each airport, accept the reduced-size model as the minimum-size model if the residuals from the training sample and the prediction errors from the testing sample pass both the t test and the Kolmogorov-Smirnov test at level 0.05, i.e., the null hypotheses of that the sample mean of squared residuals from the full-size model is greater or equal to that from the reduced-size model and they have the same distribution were not rejected.

viii. For each airport for which no model has yet been accepted, an additional factor is added.

   a. For the Generated Delay model, the variable is added in the following order: (1) add Terminal Weather first since it appeared at 16 models, (2) add a variable with highest variable importance and the value of importance is greater than 30% in the MARS output as calculated for the in full-size model, from among the variables not yet considered for addition to the model, (3) if importance of Ratio of Departure Demand and Departure Throughput in 30-minute window is less than Ratio of Departure Demand and ADR in 30-minute window or 15-minute window, then replace it with the variable with high importance, (4) add the variable which has the lowest correlation with the factors in the model. (Note: The MARS software assigns a variable importance rating to each variable which gives an indication of the predictive strength of each variable.)

   b. For Absorbed Delay model, select the variable with the highest variable importance as calculated for the in full-size model, from among the variables not yet considered for addition to the model. For reduced-size Absorbed Delay model, only one additional variable is needed to pass the hypothesis tests.

ix. Continue Steps v to viii till both tests are not rejected.

The independent variables in the reduced-size model are the significant factors, which we are interested in, for delays at corresponding airports. The variable selection process is graphically expressed in Figure **4.1**

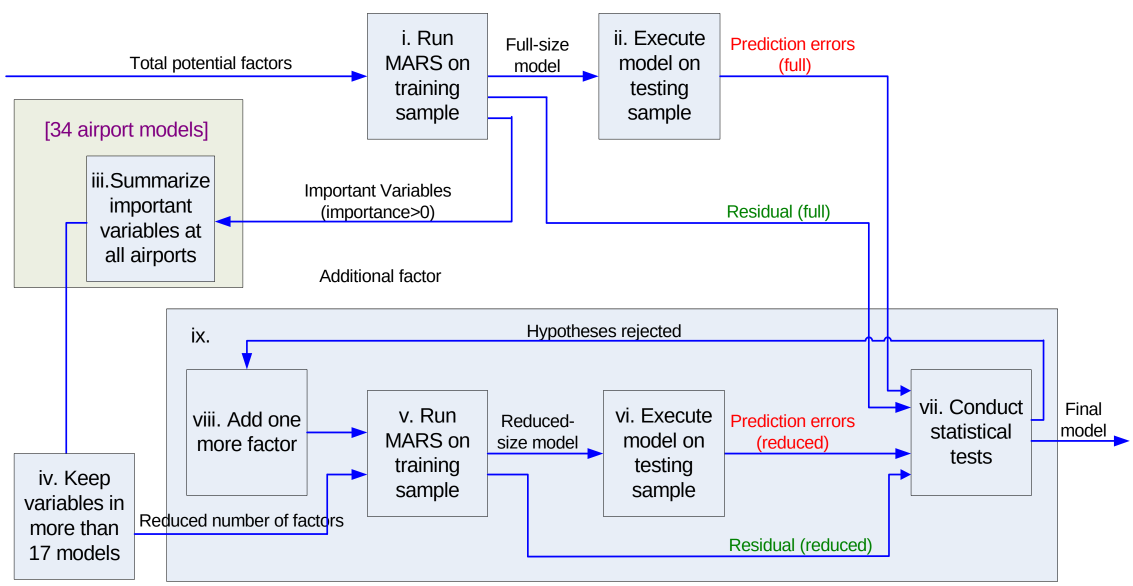


Figure **4.1**: Steps of Variable Selection in the Final Model

After these steps were completed, fewer factors for each airport were left in the final models. Paired t-tests and non-parametric Kolmogorov-Smirnov tests were conducted on the residuals of training samples from original full-size models and reduced-size final models. The same tests were also conducted on the prediction errors of the testing samples from the full-size and reduced-size models.

The test results for Airport Generated Delay models are provided in Table **A.3**. The results can neither reject the null hypotheses that the full size models have smaller squared residuals nor the null hypothesis that the squared prediction errors from both size models have the same distributions at 5% significant level at all OEP 34 airports.

The test results for Airport Absorbed Delay models are provided in Table **A.4**. For all the OEP 34 airports, the test results cannot reject the null hypotheses of equal means and same distributions. Hence, there is not strong evidence to support the hypotheses that the residuals or prediction errors from the reduced-size delay absorption models have different distributions from the full-size models.

4.2.2 Selected Factors

4.2.2.1 Factors for Airport Generated Delay

The most common variables selected for Airport Generated Delay in final models are listed in Table **4.2**. The complete list of predictors in each airport model is in Appendix Table **A.5**.

.

Table **4.2**: Selected Factors Used in Equations for Generated Delay at Each Airport (in the order of average Airport Delay per flight in summer 2005)

		Inbound	Airport			Airline			Outbound	
code	Average airport delay (min) summer 05	Inbound delay (min)	Departure rho 30 (demand/ throughput in 30min)	Departure rho ADR15 (demand/ADR in 15min)	Terminal weather	Schedule departure time	Carrier delay (min)	Swap aircraft rate	GDP holding time (min)	Actual enroute time weather
PHL	25.14		x			x	x	x	x	x
JFK	22.56	x	x			x	x	x	x	
EWR	20.67		x			x	x	x	x	
ORD	17.99	x	x		x	x	x	x	x	x
MSP	16.80	x		x	x	x	x		x	
MIA	15.89	x	x		x	x	x		x	
IAD	15.83	x	x			x	x	x	x	
IAH	15.62	x		x	x	x	x		x	
LGA	14.65	x	x		x	x	x	x	x	x
DTW	14.32	x		x		x	x		x	
CLT	14.19	x		x	x		x	x	x	
ATL	13.80	x	x			x	x	x	x	
DFW	13.26	x	x		x	x	x	x	x	
BOS	12.33	x		30-min		x	x	x	x	
PHX	11.45	x	x			x	x	x	x	
FLL	11.41	x	x			x	x	x	x	
DCA	11.09	x	x	x	x		x	x	x	x
MDW	11.05	x				x	x	x	x	
CVG	10.14	x		30-min		x	x	x	x	x
MEM	9.96	x		x	x		x	x	x	

		Inbound	Airport			Airline			Outbound	
code	Average airport delay (min) summer 05	Inbound delay (min)	Departure rho 30 (demand/ throughput in 30min)	Departure rho ADR15 (demand/AD R in 15min)	Terminal weather	Schedule departure time	Carrier delay (min)	Swap aircraft rate	GDP holding time (min)	Actual enroute time weather
BWI	9.68	x	x		x	x	x	x	x	
LAS	9.20	x	x			x	x	x	x	
CLE	9.08	x	x			x	x	x	x	
TPA	8.44	x	x			x	x	x	x	
DEN	8.39	x	x				x	x	x	
PIT	8.32	x	x		x		x	x	x	
MCO	8.30	x	x		x	x	x	x	x	
SEA	8.17	x	x			x	x	x	x	
SLC	7.76	x		x			x	x	x	
STL	6.87	x	x			x	x		x	
LAX	6.74	x	x			x	x	x	x	
SFO	6.41	x		x			x	x	x	
SAN	5.35	x	15-min			x	x	x	x	
PDX	4.21	x	x			x	x	x	x	

* 15-min and 30-min in the 4th and 5th column from left denote the ratio calculated in 15-minute and 30-minute time window which are different from other entries in the same column.

The Departure Demand Ratio reflects the airport operation and airline schedule structure, and it is considered to be a very important variable influencing the airport operation performance (Ball et al 2006). In Table **4.2**, two correlated variables about Departure Demand Ratio, Ratio of Departure Demand and Departure throughput in 30-minute time window and Ratio of Departure Demand and Airport Departure Rate (ADR) in 15-minute time window, are the selected predictors for 33 airports. The only exception is MDW airport. For MDW, the variable related to Departure Demand Ratio is not even in the full-size model.

4.2.2.2 Factors for Airport Absorbed Delay

The factors that appeared most commonly in the model for Airport Absorbed Delay at all 34 OEP airports in Table **4.3** are Inbound Delay, Scheduled Turn-around Time and Carrier Delay. Twenty nine airports have GDP Holding Time, 24 airports have Number of Seats and 7 airports have Scheduled Departure Time as additional factors. The complete list of predictors in each airport model is in Appendix Table **A.6**.

As shown in Table **4.2** and **4.3**, the selected factors for airport delays differed among airports. The next chapter will explain the final models in detail.

Table **4.3**: Selected Factors Used in Equations for Absorbed Delay at Each Airport (in the order of average Airport Delay per flight in summer 2005)

		Inbound	Airline				Outbound
Airport	Avg.airport delay(min)	Inbound Delay (min)	Sch. Dep. time	Sch. Turn time	Carrier Delay (min)	# of seats	GDP Time (min)
PHL	25.14	x	x	x	x	x	x
JFK	22.56	x		x	x		
EWR	20.67	x		x	x	x	x
ORD	17.99	x		x	x	x	x
MSP	16.80	x		x	x		x
MIA	15.89	x		x	x	x	x
IAD	15.83	x		x	x	x	x
IAH	15.62	x		x	x	x	x
LGA	14.65	x		x	x	x	x
DTW	14.32	x		x	x		x
CLT	14.19	x		x	x	x	x
ATL	13.80	x		x	x	x	x
DFW	13.26	x		x	x	x	x
BOS	12.33	x		x	x		x
PHX	11.45	x		x	x		
FLL	11.41	x	x	x	x	x	x
DCA	11.09	x		x	x	x	x
MDW	11.05	x		x	x		x
CVG	10.14	x		x	x	x	x
MEM	9.96	x		x	x	x	x
BWI	9.68	x		x	x	x	x
LAS	9.20	x	x	x	x	x	
CLE	9.08	x	x	x	x		x
TPA	8.44	x	x	x	x		x
DEN	8.39	x		x	x	x	x
PIT	8.32	x		x	x	x	x
MCO	8.30	x		x	x	x	x
SEA	8.17	x		x	x	x	x
SLC	7.76	x		x	x		x
STL	6.87	x	x	x	x	x	x
LAX	6.74	x		x	x	x	
SFO	6.41	x	x	x	x		x
SAN	5.35	x		x	x	x	
PDX	4.21	x		x	x	x	x

CHAPTER 5

FINAL MODELS

This chapter lists samples of final models of Airport Generated Delay and Absorbed Delay. The final models of all 34 OEP airports are provided in Appendix B.

5.1 Airport Generated Delay

The general form of the equation for Airport Generated Delay is illustrated by the model for ORD in Equation **5.1** . The response variable, Airport Generated Delay, is transformed by taking the square root to stabilize the variance of residuals. The same transformation was done on Airport Absorbed Delay as well.

Airport: ATL, **Generated Delay**

Basis Functions

===============

$$
\begin{aligned}
BF1 &= \max(0, GDPHoldingTime - 23.200);\\
BF2 &= \max(0, 23.200 - GDPHoldingTime);\\
BF3 &= \max(0, CarrierDelay - 11.857) * BF2;\\
BF4 &= \max(0, 11.857 - CarrierDelay) * BF2;\\
BF6 &= \max(0, 34.000 - ScheduleDepartureTime) * BF2;\\
BF7 &= \max(0, DepartureDemandRatio30 - 6.160);\\
BF8 &= \max(0, 6.160 - DepartureDemandRatio30);\\
BF9 &= \max(0, SwapAircraftRate - 0.330);\\
BF10 &= \max(0, 0.330 - SwapAircraftRate);\\
BF12 &= \max(0, 56.000 - CarrierDelay) * BF10;\\
BF13 &= \max(0, InboundDelay - 124.250) * BF10;\\
BF14 &= \max(0, 124.250 - InboundDelay) * BF10;\\
BF15 &= \max(0, ScheduleDepartureTime - 54.000);\\
BF16 &= \max(0, 54.000 - ScheduleDepartureTime);\\
BF17 &= \max(0, CarrierDelay - 3.333) * BF16;\\
BF18 &= \max(0, 3.333 - CarrierDelay) * BF16;\\
\\
\hat{g} &= \max(0, 10.785\\
&\quad + 0.041 * BF1 - 0.074 * BF2\\
&\quad + .665814E\text{-}03 * BF3 - 0.003 * BF4\\
&\quad - 0.010 * BF6\\
&\quad - 0.030 * BF7 - 0.653 * BF8\\
&\quad + 1.934 * BF9 - 1.150 * BF10\\
&\quad - 0.163 * BF12\\
&\quad - 0.092 * BF13 + 0.031 * BF14\\
&\quad + 0.027 * BF15\\
&\quad + 0.033 * BF16\\
&\quad + .961083E\text{-}03 * BF17 - 0.005 * BF18)
\end{aligned}
\qquad \textbf{5.1}
$$

Where, $\hat{g}$ is the estimate of $\sqrt{AirportGeneratedDelay}$

The model has an additive form. Each of the summands is a function of a basis function. Each basis function represents a linear contribution for a predictor over a certain range of values. For example, the two basic functions BF1 and BF2 in Airport Generated

Delay model of ATL represent two different segments of the range of values for GDP Holding Time (see Equation **5.1**). Each segment has a different coefficient in the final model: e.g., 0.041 for BF1 and -0.074 for BF2. This can be seen in the plot of Figure **5.1**. These different slopes indicate different degrees of influences of GDP Holding Time on Airport Generated Delay.

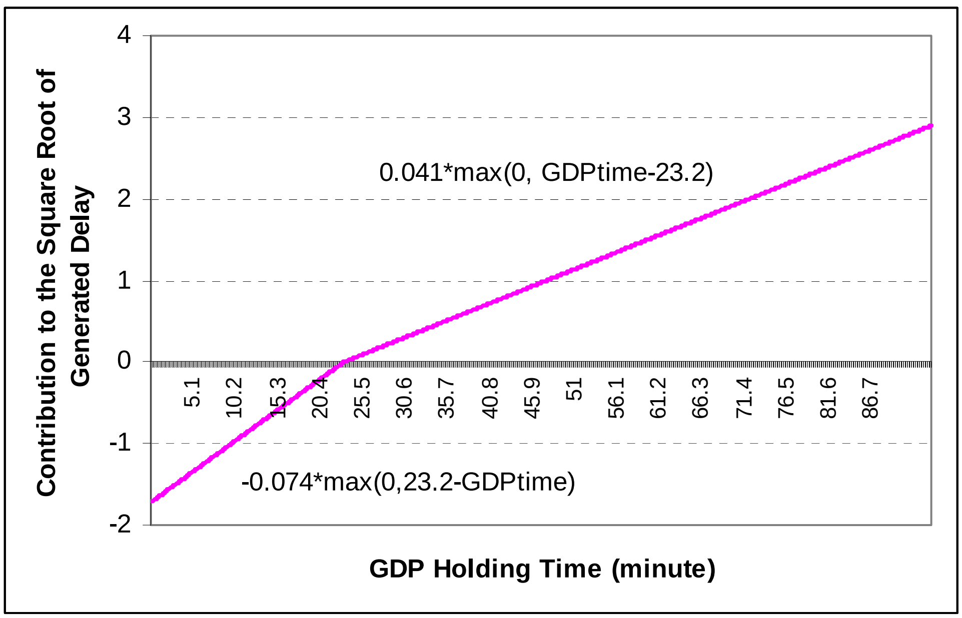


Figure **5.1**: Graphical Example of the Contributions of a Pair of Basis Functions, BF1 and BF2, to the Square Root of Generated Delay at ATL.

The model also includes interaction terms. For example, the pair of BF12 is the interaction terms of Carrier Delay and Swap Aircraft Rate. Only when Swap Aircraft Rate is less than 0.33 and Carrier Delay less than 56 minutes will BF12 have positive value. The coefficients of BF12 in final model explain the joint impact of Carrier Delay and Swap Aircraft Rate.

5.2 Airport Absorbed Delay

Similarly, the models for Airport Absorbed Delay are in the same format as Generated Delay. The difference is that the value of Absorbed Delay is negative. The transformed response variable is $-\sqrt{-AbsorbedDelay}$. Equation **5.2** shows the final Airport Absorbed Delay model of ORD.

Airport: ATL, **Absorbed Delay**

Basis Functions

===============

BF1 = max(0, InboundDelay − 18.222);
BF2 = max(0, 18.222 − InboundDelay);
BF3 = max(0, TurnaroundTime − 84.500);
BF4 = max(0, 84.500 − TurnaroundTime);
BF5 = max(0, InboundDelay − 21.000) * BF4;
BF6 = max(0, 21.000 − InboundDelay) * BF4;
BF7 = max(0, CarrierDelay − 39.667);
BF8 = max(0, 39.667 − CarrierDelay);
BF10 = max(0, 54.700 − GDPHoldingTime);
BF11 = max(0, NumberSeats − 161.400);
BF12 = max(0, 161.400 − NumberSeats);
BF13 = max(0, InboundDelay − 69.333) * BF10;
BF14 = max(0, 69.333 − InboundDelay) * BF10;
BF15 = max(0, InboundDelay − 8.750) * BF8; **5.2**
BF16 = max(0, 8.750 − InboundDelay) * BF8;
BF17 = max(0, InboundDelay + .441525E-06) * BF12;
BF18 = max(0, - .441525E-06 − InboundDelay) * BF12;

â = min(0, -1.324
− 0.027 * BF1 + 0.053 * BF2
+ 0.006 * BF3 + 0.038 * BF4
+ .480074E-03 * BF5 − 0.001 * BF6
+ 0.004 * BF7 − 0.033 * BF8
− 0.025 * BF10
+ 0.013 * BF11 + 0.002 * BF12
+ .596830E-03 * BF13 + .154085E-03 * BF14
+ .208147E-03 * BF15 + 0.001 * BF16
− .253714E-03 * BF17 − .252755E-03 * BF18)

where, â is the estimate of $-\sqrt{-\text{AirportAbsorbedDelay}}$

5.3 Airport Delay

Airport Delay is the summation of Airport Generated Delay and Absorbed Delay. Since the outputs of both the Generated Delay models and the Absorbed Delay models are the square roots of delays, each point estimate provided by these models needs to be transformed back to its original scale in minute so that the point estimate of Airport Delay can be calculated (see Equation **5.3**).

$$\text{Airport }\hat{\text{D}}\text{elay}_{epoch} = \hat{g}_{epoch}{}^{2} - \hat{a}_{epoch}{}^{2}$$

where, $\hat{g}_{epoch}$ is the estimate of $\sqrt{\text{AirportGeneratedDelay}}$ for a 15-minute epoch **5.3**

$\hat{a}_{epoch}$ is the estimate of $-\sqrt{-\text{AirportAbsorbedDelay}}$ for a 15-minute epoch

The models developed can not only provide the point estimate of the mean of Airport Delay using the Equation **5.3**, but the prediction intervals for given values of the predictors. The way to calculate an approximate standard deviation of back-transformed Airport Delay (Equation **5.4**) is described as follows.

Let $G_i = \sqrt{\text{GeneratedDelay}_i}$, $\widehat{E[G_i]} = \hat{g}_i$ and $\widehat{V[G_i]} = \hat{\sigma}_g{}^2 \approx \sum_n (\hat{g}_i - g_i)^2 \Big/ (n - p_g)$.

Let $A_i = -\sqrt{-\text{AbsorbedDelay}_i}$, $\widehat{E[A_i]} = \hat{a}_i$ and $\widehat{V[A_i]} = \hat{\sigma}_a{}^2 \approx \sum_n (\hat{a}_i - a_i)^2 \Big/ (n - p_a)$,

where p is number of basis functions plus 1 (intercept),
and n is number of cases.

Let Airport Delay = GeneratedDelay + AbsorbedDelay

$$T_i = G_i^2 - A_i^2,$$

and

$$V[T_i] = V\left[G_i^2\right] + V\left[A_i^2\right] - 2Cov\left[G_i^2, A_i^2\right].$$

Based on a first-order Taylor series expansion,

$$f(Y) - f(\mu_Y) \approx (Y - \mu_Y) f'(\mu_Y),$$

we get

$$\begin{aligned}
&E\big(f(Y)\big) \approx f(\mu_Y), \\
&V\big[f(Y)\big] \approx V(Y)\big[f'(\mu)\big]^2, \\
&\text{and } Cov\big[f(Y), f(Z)\big] \approx Cov(Y, Z) f'(\mu_Y) f'(\mu_Z).
\end{aligned}$$

It follows that

$$\begin{aligned}
&V\big[G_i^2\big] \approx \widehat{V[G_i]}(2\hat{g}_i)^2 = 4\hat{\sigma}_g{}^2\hat{g}_i{}^2, \\
&V\big[A_i^2\big] \approx \widehat{V[A_i]}(2\hat{a}_i)^2 = 4\hat{\sigma}_a{}^2\hat{a}_i{}^2, \\
&\text{and } Cov\big[G_i^2, A_i^2\big] \approx 4\hat{g}_i\hat{a}_i\widehat{Cov(G_i, A_i)}.
\end{aligned} \qquad \mathbf{5.4}$$

Overall, we have that

$$V[T_i] \approx 4\sigma_g{}^2\hat{g}_i{}^2 + 4\sigma_a{}^2\hat{a}_i{}^2 - 8\hat{g}_i\hat{a}_i\widehat{Cov(G_i, A_i)}.$$

5.4 Results

An inspection of models in Equation **5.1** to Equation **5.2** of ATL and all 34 OEP airport models in Appendix B shows that no two airports share the same model for either Generated Delay or Absorbed Delay. Different airport models have different factors, different knots over the factor's value space, and different coefficients. The models are unique for each airport. Hence, using one general model to predict delays at all airports is questionable, at least to the level of fidelity in this research (15-minute increments).

CHAPTER 6

MODEL VALIDATION

The models derived in this research are used to predict airport performance at a future time. It is of paramount importance to validate the predictions from the regression models since regression analysis is susceptible to overfitting the training data.

There are three ways to examine model validity (Steyerberg et al. 2000).

- Apparent validation tests the regression model using the same data that was used to develop the model. This type of validation is mainly for assessing model assumptions via residuals.
- Internal validation tests the regression model by using the data coming from the same underlying population.
- External validation tests the model by using the data coming from a related but slightly different population.

A good regression model should be able to estimate the data out of the training sample without suspiciously high error measures. Both internal and external validations evaluate model performance via prediction errors. External validation is easier to calculate than the internal one because external validation just needs a new data sample different from training sample, while the internal validation requires another random sample from same population.

In this research, apparent validations were conducted during the model building process, and the results are shown in Appendix Figure **A.1** to **A.8**. This chapter only show the external validation on the models derived for airports.

6.1 External Validation

A hold-out sample from the last 15 days of August 2005 and another data sample from the last 15 days of August 2006 (August 31^{st} was excluded from both years) were used to measure the prediction accuracy for external validity. External validation more accurately measures the prediction performance of the models.

Figure **6.1** compares the actual Airport Delay and model output at each epoch on one day, August 24 2005 at ATL.

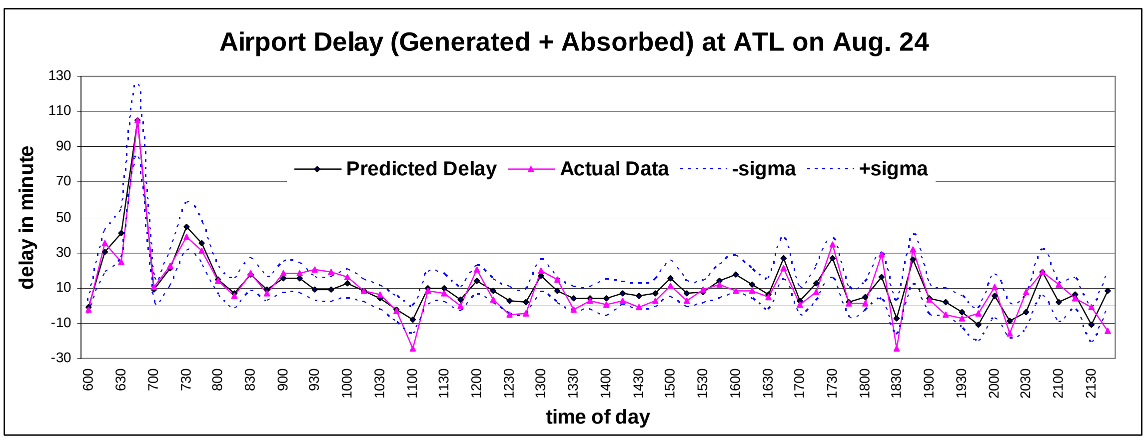


Figure **6.1**: Comparison of Estimated and Actual Airport Delay at ORD on Aug. 24, 2005

The dots in light color represent the estimated Airport Delays for each 15-minute epoch given the values of predictors. The dots in dark color are the actual delay for each epoch. The dashed lines are the square of estimated Airport Delay plus or minus one estimated standard deviation of the error term of Airport Delay based on Equation **5.4**.

Not all the predictions for epochs on August 24, 2005 at ATL fall within $\hat{\sigma}$ of the regression line. Table **6.1** summarized the percentage of actual Airport Delay data within $1\hat{\sigma}$ and $2\hat{\sigma}$ of the regression line. About 77% of validation data of 2005 and 73.2% of validation data of 2006 fall within $\hat{\sigma}$ of the regression line.

Table **6.1**: Summary of the Percentage of Actual Airport Delay Data (w) falls within 68% and 95% prediction intervals for 34 OEP Airports.

		Minimum percentage	Average percentage
2005	Percentage of actual data in $\hat{w} \pm \hat{\sigma}$	67.9%	77.0%
	Percentage of actual data in $\hat{w} \pm 2\hat{\sigma}$	90.6%	94.9%
2006	Percentage of actual data in $\hat{w} \pm \hat{\sigma}$	67.0%	73.2%
	Percentage of actual data in $\hat{w} \pm 2\hat{\sigma}$	90.2%	93.6%

Hence, the models perform very well on both 2005 and 2006 data, while 2005 is a little better than 2006. The detailed information about each airport is in Table **A.7**.

Several general validation statistics are provided in the next section.

6.1.1 Validation Statistics

Several measures of the overall estimation performance of the models for Generated and Absorbed Delay are calculated to assess how well the model fits the data. These are the mean transformed prediction error,

$$MTPE = \frac{1}{m}\sum\left(\text{estimate}^2 - \text{actual Data}\right),$$

and the mean absolute transformed prediction error,

$$MTAPE = \frac{1}{m}\sum\left|\text{estimate}^2 - \text{actual Data}\right|,$$

where, m is the number of records in the validation data set, the second half of August.

The direct output of the fixed models are $\sqrt{\text{Airport Generated Delay}}$ and $-\sqrt{-\text{Airport Absorbed Delay}}$. Each model prediction was transformed back to the original scale to obtain a prediction of Airport Generated and Absorbed Delay. This enables the calculation of errors in the same units and scales as the actual delays.

The MTPE is obtained by subtracting the actual delays from the transformed predictions, and then averaging them. This statistic takes account of the sign of positive and negative errors. It measures the bias of predictions from the fixed model. The ideal MTPE is zero.

The MATPE is obtained by summing the absolute prediction errors for delays. Since there are both negative and positive errors, taking the absolute value of errors estimates the average magnitude of the errors without considering their direction. This statistic measures the variability of the predictions from the fixed model.

6.1.2 Validation Results for 2005

The values of validation statistics defined in the previous subsection for the overall performance are given in Table **6.2**. The airports are listed in the order of average Airport Delay per flight in the summer 2005.

Table **6.2**: Validation Results of Airport Delays at 34 OEP Airports of Aug.2005

	Generated Delay			Absorbed Delay			Airport Delay			
Airport	R^2	MTPE	MATPE	R^2	MTPE	MATPE	Actual Mean	MTPE	Actual Std	MATPE
PHL	0.58	1.4	8.7	0.51	1.0	1.8	16.2	2.4	17.4	9.2
JFK	0.60	-0.4	8.0	0.54	1.3	2.5	17.3	0.9	21.4	8.7
EWR	0.52	0.6	7.7	0.49	1.2	2.5	14.7	1.8	18.3	8.5
ORD	0.72	1.7	5.5	0.52	1.0	1.8	11.2	2.7	14.6	6.3
MSP	0.62	0.9	5.9	0.59	0.2	2.3	14.1	1.1	21.3	6.5
MIA	0.67	-1.0	6.1	0.42	0.7	2.1	9.2	-0.2	20.3	7.2
IAD	0.53	1.2	7.2	0.42	1.4	2.5	8.7	2.6	21.0	8.3
IAH	0.65	-2.2	8.3	0.46	1.2	2.1	12.8	-1.0	22.6	9.2
LGA	0.49	-2.2	7.3	0.59	1.0	2.3	11.8	-1.2	14.8	7.8
DTW	0.71	0.5	5.2	0.51	0.9	2.5	13.0	1.4	21.1	6.1
CLT	0.68	-1.3	5.6	0.47	1.2	2.2	9.3	-0.1	17.4	6.3
ATL	0.68	-1.5	5.8	0.60	0.7	2.3	12.5	-0.8	16.7	6.6
DFW	0.70	-1.0	5.4	0.41	0.9	1.8	10.0	0.0	15.3	6.1
BOS	0.64	-0.4	5.3	0.47	1.2	2.4	7.1	0.7	13.1	6.3
PHX	0.62	-0.5	4.8	0.58	0.6	2.1	7.6	0.0	14.9	5.9
FLL	0.60	-0.2	5.3	0.47	0.9	1.9	7.0	0.7	20.0	6.3
DCA	0.63	-0.1	3.9	0.47	1.3	2.1	5.5	1.2	12.9	4.7
MDW	0.65	0.4	4.2	0.47	1.0	1.8	5.6	1.3	13.6	5.1
CVG	0.69	-1.3	4.8	0.58	1.1	2.3	6.9	-0.2	14.7	5.6
MEM	0.69	-0.9	5.8	0.51	1.3	2.6	8.5	0.4	20.2	7.0
BWI	0.60	-0.2	4.8	0.38	1.0	1.8	7.6	0.8	16.5	5.7
LAS	0.63	-0.1	3.9	0.46	0.8	1.8	5.9	0.8	12.3	4.7
CLE	0.68	-1.8	4.9	0.52	1.0	2.2	5.4	-0.9	20.6	5.9
TPA	0.66	-0.8	4.8	0.47	1.2	2.1	5.2	0.4	19.7	6.0
DEN	0.61	-0.2	3.9	0.50	1.0	2.0	4.3	0.8	10.9	4.9
PIT	0.69	-1.0	4.1	0.46	1.4	2.4	2.7	0.4	17.8	5.4
MCO	0.68	-0.7	4.5	0.39	0.8	2.0	6.0	0.1	16.5	5.4
SEA	0.66	-0.6	3.6	0.66	0.9	2.2	5.0	0.3	14.5	4.5
SLC	0.58	-1.4	4.4	0.51	0.9	1.9	6.1	-0.5	15.3	5.2
STL	0.67	-0.2	4.3	0.44	1.2	2.1	3.8	1.0	11.6	5.1
LAX	0.57	-0.2	3.4	0.50	0.5	1.8	5.0	0.3	9.8	4.2
SFO	0.59	-0.3	4.5	0.65	1.0	2.4	3.5	0.7	15.7	5.7
SAN	0.50	-0.8	4.8	0.46	1.0	2.1	3.6	0.1	13.9	5.7
PDX	0.63	-1.6	3.4	0.64	0.7	2.1	1.8	-0.8	12.9	4.5
average	**0.63**	**-0.5**	**5.3**	**0.50**	**1.0**	**2.1**	**8.1**	**0.5**	**16.5**	**6.2**
min	**0.49**	**-2.2**	**3.4**	**0.38**	**0.2**	**1.8**	**1.8**	**-1.2**	**9.8**	**4.2**
max	**0.72**	**1.7**	**8.7**	**0.66**	**1.4**	**2.6**	**17.3**	**2.7**	**22.6**	**9.2**

As can be seen from Table **6.2**, the MTPE from validation samples in 2005, Aug. 1st to Aug. 15th, are between -2.2 minutes at IAH and LGA and 1.7 minutes at ORD for Airport Generated Delay; and between 0.2 minutes at MSP and 1.4 minutes at PIT for Absorbed Delay. The average value of MATPE for Airport Generated Delay is 5.3 minutes; the average value of MATPE for Airport Absorbed Delay is 2.1 minutes. The standard deviation of Airport Generated Delay of all 34 airports in summer 2005 is 20.0 minutes. Based on the comments from experts of aviation modeling, prediction errors up to 15 minutes are acceptable. Therefore, the range of variation of prediction errors is reasonable. Predictions for the 2005 validation data correspond reasonably well to the actual observations.

6.1.3 Validation Results for 2006

Compared to the 2005 validation data, the August 2006 data is more likely to be different from the 2005 training data. After one year, airlines may have changed their operation strategies and airports may have improved operational conditions. The validation conducted on 2006 data provides a stronger test of model prediction performance (see Table **6.3**).

Table **6.3**: Validation Results of Airport Delays at 34 OEP Airports of Aug.2006

Airport	Generated Delay			Absorbed Delay			Airport Delay			
	R^2	MTPE	MATPE	R^2	MTPE	MATPE	Actual Mean	MTPE	Actual Std	MATPE
PHL	0.58	-0.6	11.8	0.51	1.2	2.1	19.4	0.6	26.6	12.4
JFK	0.60	-5.5	11.3	0.54	1.0	3.0	23.7	-4.5	26.0	11.9
EWR	0.52	-4.9	11.4	0.49	1.1	2.5	21.2	-3.8	21.1	11.8
ORD	0.72	1.8	7.5	0.52	1.5	2.4	13.1	3.4	16.8	8.5
MSP	0.62	0.8	6.0	0.59	1.6	2.8	8.6	2.4	18.5	7.5
MIA	0.67	-0.2	8.0	0.42	1.1	2.6	10.4	0.9	26.2	9.6
IAD	0.53	1.3	8.0	0.42	1.8	2.7	9.1	3.2	22.1	9.5
IAH	0.65	-0.8	7.8	0.46	1.0	2.3	12.2	0.1	20.9	8.8
LGA	0.49	-1.7	8.1	0.59	1.1	2.6	11.6	-0.6	15.5	8.9
DTW	0.71	0.7	4.7	0.51	1.3	2.3	7.4	2.1	19.2	5.8
CLT	0.68	-2.0	7.2	0.47	1.5	2.6	10.1	-0.5	21.1	8.3
ATL	0.68	-3.4	6.7	0.60	0.8	2.0	15.3	-2.6	16.3	7.2
DFW	0.70	0.4	6.2	0.41	1.0	1.8	10.8	1.4	15.3	7.0
BOS	0.64	-0.1	5.6	0.47	1.1	2.5	8.2	1.0	15.7	6.6
PHX	0.62	-1.6	7.4	0.58	0.7	2.0	10.2	-0.9	18.8	8.1
FLL	0.60	1.2	5.2	0.47	2.0	2.6	3.6	3.2	15.1	6.9
DCA	0.63	-0.3	5.3	0.47	1.6	2.4	8.3	1.2	18.2	6.0
MDW	0.65	0.0	5.8	0.47	0.9	1.6	9.8	0.9	16.2	6.4
CVG	0.69	1.4	5.3	0.58	1.3	2.2	6.1	2.6	22.2	6.5
MEM	0.69	-0.2	4.9	0.51	1.9	3.0	5.4	1.8	20.4	6.3
BWI	0.60	0.4	4.3	0.38	1.2	2.0	6.0	1.6	14.3	5.2
LAS	0.63	-2.7	5.1	0.46	0.2	2.0	10.3	-2.5	13.5	5.8
CLE	0.68	0.1	5.3	0.52	1.3	2.7	6.8	1.4	25.5	6.7
TPA	0.66	-2.5	5.4	0.47	1.3	2.2	5.4	-1.2	17.4	6.5
DEN	0.61	0.8	4.7	0.50	1.2	2.0	5.5	2.0	13.7	5.8
PIT	0.69	-2.3	5.4	0.46	0.7	2.4	6.6	-1.7	16.5	6.2
MCO	0.68	-1.9	5.2	0.39	0.6	1.9	7.8	-1.3	18.8	5.9
SEA	0.66	-2.3	4.5	0.66	1.2	2.1	8.1	-1.1	15.2	5.1
SLC	0.58	-2.4	4.6	0.51	1.4	2.4	5.7	-1.0	13.5	5.7
STL	0.67	-0.1	3.1	0.44	1.6	2.2	3.4	1.5	13.5	4.4
LAX	0.57	-2.1	4.1	0.50	0.5	2.1	7.1	-1.6	10.8	4.6
SFO	0.59	-0.8	4.4	0.65	1.2	2.8	2.9	0.4	13.2	5.7
SAN	0.50	-0.8	3.8	0.46	1.2	2.3	3.2	0.4	10.9	4.9
PDX	0.63	-1.2	3.3	0.64	1.1	2.1	3.4	-0.2	14.8	4.1
average	**0.63**	**-0.9**	**6.1**	**0.50**	**1.2**	**2.3**	**9.0**	**0.3**	**17.8**	**7.1**
min	**0.49**	**-5.5**	**3.1**	**0.38**	**0.2**	**1.6**	**2.9**	**-4.5**	**10.8**	**4.1**
max	**0.72**	**1.8**	**11.8**	**0.66**	**2.0**	**3.0**	**23.7**	**3.4**	**26.6**	**12.4**

The MATPEs from 2006 validation samples are between 3.1 minutes at STL and 11.8 minutes at PHL for Airport Generated Delay; and between 1.6 minutes at MDW and 3.0 minutes at JFK for Absorbed Delay. Overall, for Airport Delay, the MATPEs are from 4.1 minutes at PDX and 12.4 minutes at PHL.

6.1.4 Comparison of Validation Results of 2005 and 2006

For most airports, the validation results for August 2005 are slightly better than 2006. This is to be expected since the airport performance in August 2005 should be more similar to June and July 2005 than August 2006. Figure **6.2** shows the histograms of MTPE from 2005 and 2006 together, so that the difference can be seen easily.

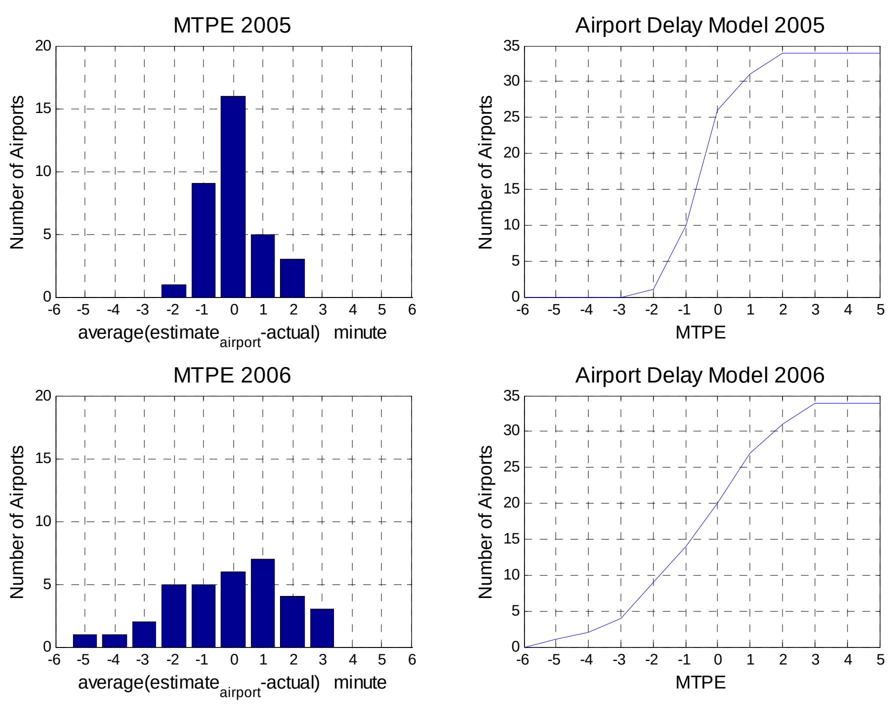


Figure **6.2**: Comparison of MTPEs of Airport Generated Delay Using Validation Data from 2005 and 2006

The MTPEs of Generated Delay at JFK, EWR, ATL, and LAS in August 2006 are outside the left side boundary of all MTPEs in 2005; and the MTPE of ORD, IAD, FLL, and CVG in 2006 are outside the right side boundary of all MTPE in 2005. The MATPEs at PHL, JFK, and EWR in 2006 are outside the right boundary of all MATPEs in 2005 in Figure **6.3**. It also shows that the models perform better for data in 2005 than 2006.

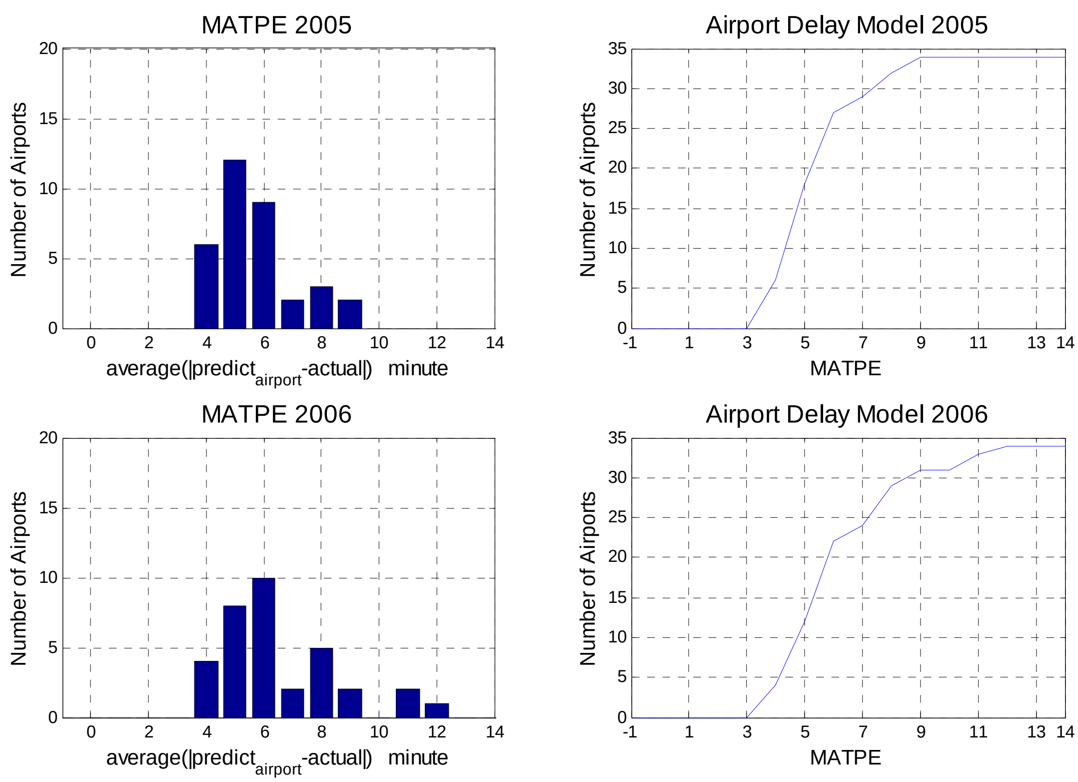


Figure **6.3**: Comparison of MATPEs of Airport Absorbed Delay Using Validation Data from 2005 and 2006

All MATPEs for 2005 and 2006 are within 15 minutes. However, the predictions for August 2005 are slightly better than 2006. These results reveal the fact that the operations of NAS keeps changing, and the models need to be updated in order to be able to represent the real system. For example, accounting for adaptation by using a running

60-day window to predict next 30 days might be a better strategy for using this model. In this research, we only consider summer season. The seasonal effects may also need to be included in the model when it is used to predict delays at different season.

CHAPTER 7

SENSITIVITY ANALYSIS

In addition to developing a methodology for predicting airport delay, another main purpose of this research is to provide a quantitative measurement of crucial factors' influence on airport delay. This chapter describes a method for approximate sensitivity analysis in non-linear models. The results from sensitivity analysis are also discussed.

7.1 Approach to Sensitivity Analysis

If a model is linear and the factors in the model are independent, the coefficients in the model provide a measure of the effects of the factors. However, in the multi-factor piecewise models obtained in this research, separate slopes were fitted to the observations in different regions of the predictor variables' value spaces. The interaction terms are context dependent as well. Therefore, the coefficient of a predictor is not directly interpretable as a global, context-independent measure of the factor's impact. An approximate approach has to be designed to measure the degree of influence of predictors in the non-linear model.

7.1.1 General Equation for Sensitivity Analysis

The sensitivity analysis is conducted by adjusting one predictor while holding other predictors constant. Let $y_i = f(x_i, \mathbf{x}_j)$, $E(y)_i = E\left(f(x_i, \mathbf{x}_j)\right)$, and $y_i + \Delta_y = f(x_i + \Delta_x, \mathbf{x}_j)$ where $i \neq j$. For a non-linear model, given the same Δ_x, different x_i will yield different Δ_y; however, $E\left(y_i + \Delta_y\right) = E\left(f(x_i + \Delta_x, \mathbf{x}_j)\right)$ holds, hence

$$\begin{aligned} E\left(\Delta_y\right) &= E\left(f(x_i + \Delta_x, \mathbf{x}_j)\right) - E\left(y_i\right) \\ &= E\left(f(x_i + \Delta_x, \mathbf{x}_j)\right) - E\left(f(x_i, \mathbf{x}_j)\right) \end{aligned}. \qquad \textbf{7.1}$$

The expected difference between the expected prediction from the adjusted data set and the expected prediction from the original data set is the approximate results of changes of x_i alone, which reflects the unique contribution of that predictor to the response variable, if that predictor's value is artificially changed by a certain amount while holding the value of other predictors constant.

Since each delay model was developed to fit the training data and the models perform best on the training data in the form of mean squared residuals, the training data was used as the original data set in this chapter.

The predictors in the models are measured in different ways. In order to compare the relative importance of these incommensurate predictors, a standardized adjustment amount was set as the percentage of the expected values of the predictor of interest across all airports in the training sample, that is

$$\Delta_x = w\%E(x_i) \qquad \textbf{7.2}$$

where, $w \in \{-50,-40,-30,-20,-10,10,20,30,40,50\}$. This is a general rule for the sensitivity analysis conducted in this research. There are specific rules for each factor to accommodate their different distribution and practical reality. These rules are described individually in detail in the following subsections.

7.1.2 Factors in Sensitivity Analysis and Their Distribution

The factors compared in the sensitivity analysis are the most common predictors in the final models, i.e., these factors are the predictors in more than 17 airport models. The factors compared for Airport Generated Delay are the GDP Holding Time, Carrier Delay, Inbound Delay, ratio of departure demand and departure throughput (30-minute window), and Swap Aircraft Rate. The factors compared for Airport Absorbed Delay are the GDP Holding Time, Carrier Delay, Inbound Delay, Scheduled Turn-around Time on Airport Absorbed Delay, and the Number of Seats.

Time of day is not included in the sensitivity analysis since airlines schedule their flight based on their market and it is solely an airline's decision about when to fly the flight. Arbitrarily adjusting the Scheduled Departure Time implies arbitrarily adjusting the airlines' schedule, and it is not practically reasonable.

Figure **7.1** plots the distributions of these selected factors at all 34 OEP airport in June and July 2005.

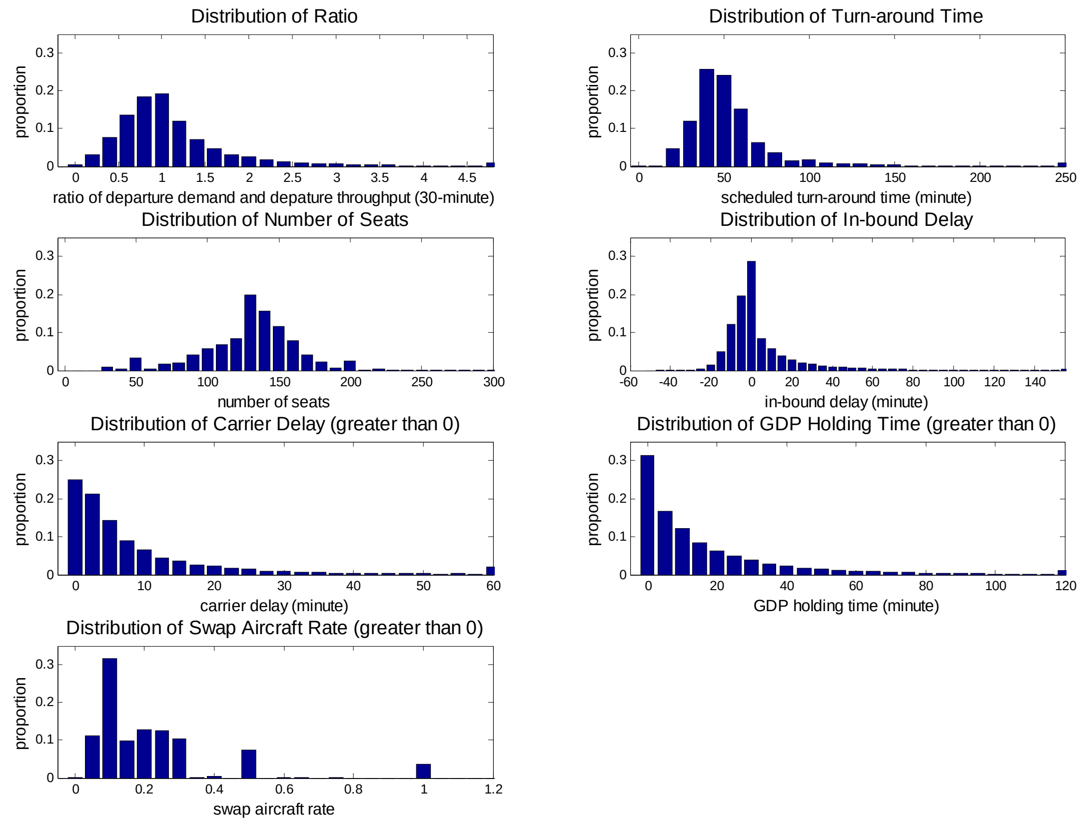


Figure **7.1**: Distribution of Value of Selected Factor at 34 OEP Airports in June and July 2005. The y-axis is the proportion of 15-minute epochs associated with the value in the x-axis.

The expected value of factors Departure Demand Ratio, Scheduled Turn-around Time, and Number of Seats was computed as the sample mean of data of June and July 2005 (training sample) from 34 airports.

Equation **7.1** provides a general rule to conduct the sensitivity analysis of predictors. However, the purpose of the sensitivity analysis conducted in this research is not solely about a factor's impact shown in the model but about its impact on the overall airport delays which can be estimated by the model given the all other conditions are the same as June and July 2005. In this research, the response variables are delays, and it is assumed that the only associated consequence of the changes of predictors in the models is delay, and the cost associated with delays is linearly correlated to the length of delay.

Suppose there was a factor in a model which has very strong statistical significance on the Airport Delays. However, its impact on the overall airport performance is very small because the bad condition of this factor does not occur very often. Putting much effort to amend that factor may only achieve a very small improvement of the overall system performance. Hence, when a factor is artificially manipulated, it is done by considering its actual distribution and within its reasonable value space.

The histograms of Carrier Delay and GDP Holding Time and Swap Aircraft Rate in Figure **7.1** plot the distribution of greater than zero data. Summarizing the data of all 34 airports in June and July 2005, only 23.1% 15-minute epochs had greater than zero GDP Holding Time, 32.4% epochs had greater than zero Carrier Delay and 4.1% epochs had greater than zero Swap Aircraft Rate value.

Using the Carrier Delay as an example, fewer than 40% of the epochs have flights with positive Carrier Delay. Imposing a certain amount of Carrier Delay to all epochs is not a realistic assumption for any airport. Furthermore, it is not right to subtract a certain amount of Carrier Delay from all epochs because this would result in more than 60% epochs having negative Carrier Delay. A negative value of Carrier Delay violates the definition of Carrier Delay in the BTS database. Hence, the adjustment made on the Carrier Delay actually extends or shrinks existing positive Carrier Delays in the Airport

Delay database, the expected value, $E(x_i)$ in Equation **7.2**, for these factors is calculated as $\overline{x}$ where $x_i > 0$, and the minimum value of Carrier Delay after adjustment is zero.

The expected value of Inbound Delay is also calculated as the sample mean of greater than zero Inbound Delays. This choice was made because the Inbound Delay has both negative and positive values in the database, and only the positive Inbound Delay is of concern.

7.1.3 Steps of Sensitivity Analysis

An approximation to the effect of manipulating one factor on the airport delays was calculated as follows:

1. Execute the regression models on the 34 original training sets and record the predictions of the squared transformed Airport Generated Delay, $\hat{g}_{origin_i}$, and Absorbed Delay, $\hat{a}_{origin_i}$
2. For each predictor:
    a. Calculate the expected value of a specific factor from all 34 OEP airports.
    b. For each record in the training set, add *w*% of the expected value of the factor to obtain an adjusted data sample, where $w \in \{-50,-40,-30,-20,-10,10,20,30,40,50\}$. If a factor is not in the model of an airport, its reduced percentage is zero. For n

factors in the sensitivity analysis, there are nx10 sets of adjusted data samples. Detailed adjustment for each factor is presented in the subsection 7.2 respectively.

c. Execute regression models on 34 adjusted data sets and record the predictions of the squared transformed Airport Generated Delay, $\hat{g}_{adjust_i}$ and Absorbed Delay, $\hat{a}_{adjust_i}$.

d. Calculate the difference between the mean of predicted Airport Generated Delay from the adjusted data set and the mean from the original data set. Both of them were transformed back to their original scale using Equation **7.3**. The same calculations were also carried out for Airport Absorbed Delay.

$$\begin{aligned}
&\text{Generated Delay Variation} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{g}_{adjust_i}{}^2 - \hat{g}_{origin_i}{}^2\right) \\
&\text{Absorbed Delay Variation} = \frac{1}{n}\sum_{i=1}^{n}\left(-\hat{a}_{adjust_i}{}^2 + \hat{a}_{origin_i}{}^2\right) \\
&\text{where, } \hat{g}_{origin} - \text{estimate of Generated Delay from original training data} \\
&\qquad \hat{g}_{adjust} - \text{estimate of Generated Delay from adjusted data} \\
&\qquad \hat{a}_{origin_i} - \text{estimate of Absorbed Delay from original training data} \\
&\qquad \hat{a}_{adjust_i} - \text{estimate of Absorbed Delay from adjusted data} \\
&\qquad\quad n - \text{number of cases in training data}
\end{aligned} \qquad \textbf{7.3}$$

3. Rank the predictor variables impact based on the slopes of delay variation value vs. increments of factor.

7.2 Sensitivity Analysis of Individual Factors

The sensitivity analysis conducted in this research provides a quantitative measure of how manipulating **one** factor affects airport performance assuming that the model accurately reflects the contribution of each factor on delay when other factors are held constant. The formula to calculate the mean and adjustment of each factor are provided in the following subsections along with the sensitivity analysis results.

7.2.1 Carrier Delay

In the Airport Delay database of summer 2005, there are only 32.4% epochs with greater than zero Carrier Delay. For the reasons explained previously, the adjusted data consists of just the records having positive Carrier Delay, and the remaining data is kept same.

The mean of Carrier Delay is the aggregated value of positive Carrier Delay in the database (Equation **7.4**). Summing over the epochs from 24 to 87 in June and July in 2005 at 34 airports obtain the total Carrier Delay. I_{epoch} is an indicator variable at an epoch. If the Carrier Delay in that epoch is positive, $I_{epoch} = 1$; otherwise, it is zero. Summing all *I*s obtains the total number of epochs having positive Carrier Delay. The average value of Carrier Delay for each of these epochs is 10.35 minutes.

$$\text{Mean Carrier Delay} = \frac{\sum_{airport=1}^{34}\sum_{day=1}^{61}\sum_{epoch=24}^{87}\text{Carrier Delay}_{airport,day,epoch}}{\sum_{airport=1}^{34}\sum_{day=1}^{61}\sum_{epoch=24}^{87} I_{airport,day,epoch}}$$

$$= 10.35(\text{minute}) \qquad \textbf{7.4}$$

where, $\text{Carrier Delay}_{\text{airport,day,epoch}}$ is average Carrier Delay at an epoch in a day at an airport

$$I_{epoch} = \begin{cases} 1, & \text{Carrier Delay}_{\text{airport,day,epoch}} > 0 \\ 0, & \text{otherwise} \end{cases}$$

Equation **7.5** formulates the approach to adjusting values of Carrier Delay. Only positive Carrier Delays are manipulated, and the adjusted value is constrained to be non-negative.

$$\text{adjusted Carrier Delay}_{\text{airport,day,epoch}}$$
$$= \begin{cases} \max\left(\text{Carrier Delay}_{\text{airport,day,epoch}} + w\% * \text{Mean Carrier Delay},\ 0\right), \\ \qquad\qquad\qquad\qquad \text{Carrier Delay}_{\text{airport,day,epoch}} > 0 \\ 0,\ \text{otherwise} \end{cases} \qquad \textbf{7.5}$$

where, $w \in \{-50,-40,-30,-20,-10,10,20,30,40,50\}$

Delay Variation at 34 OEP airports

Using Equation **7.3**, delay variations were calculated for the adjustment at each percentage for Airport Generated Delay and Absorbed Delay. Figure **7.2** plots the Variations of Airport Generated Delay vs. Adjusted portion of mean at 34 OEP airports. The x-axes are the number from -50% to 50%, they are *w*% in Equation **7.5**. Note: not all records were adjusted. Figure **7.3** plots the Variations of Airport Absorbed Delay vs. Adjusted portion of mean at 34 OEP airports.

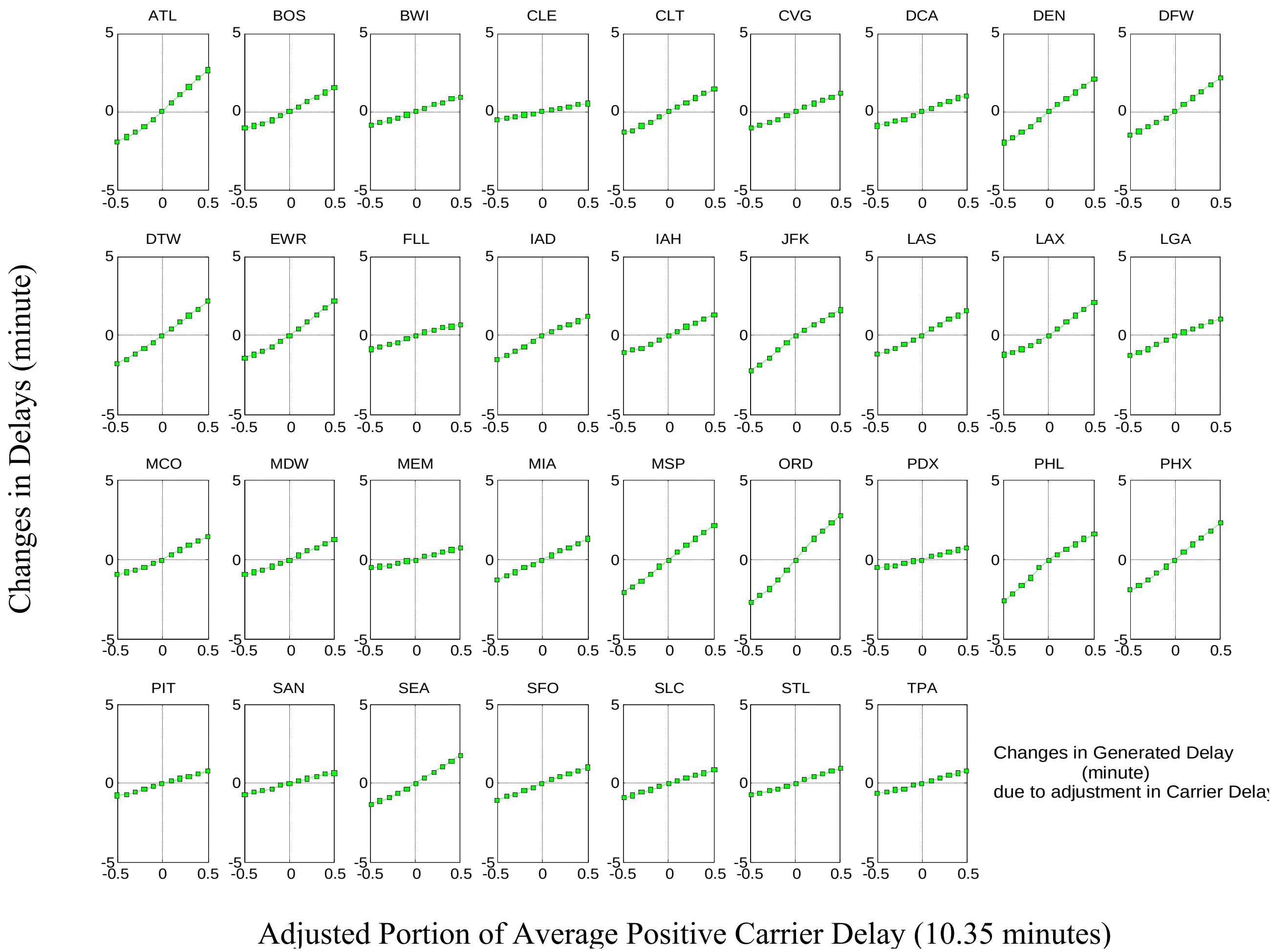


Figure **7.2**: Changes in Airport Generated Delay from Adjustment of Carrier Delay (minute)

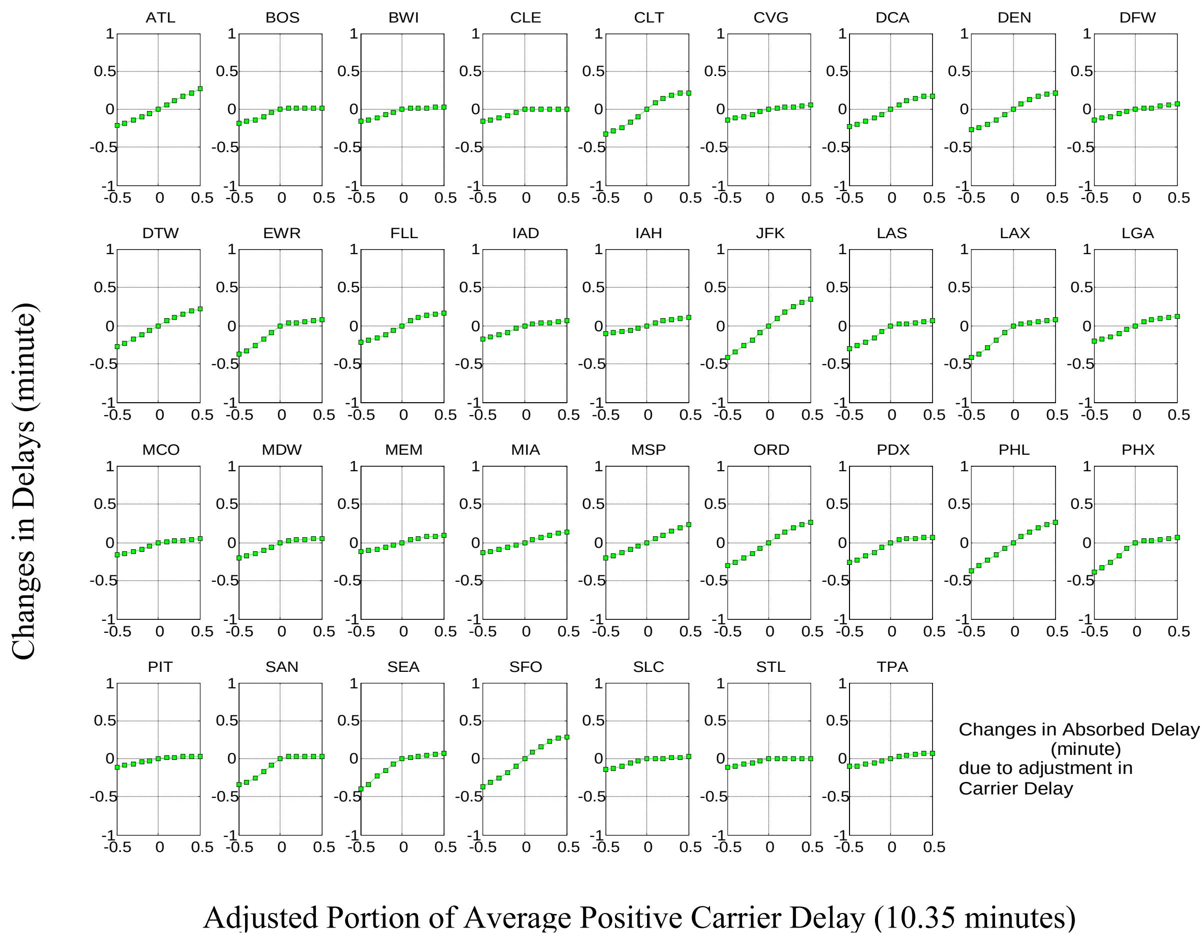


Figure **7.3**: Changes in Airport Absorbed Delay from Adjustment of Carrier Delay (minute)

In the top left panel of Figure **7.2** is the plot of the change in Generated Delay vs. increments of Carrier Delay at ATL. The plot shows that Generated Delay increases as the adjustment in the mean increases. This means that longer Carrier Delay results in longer Generated Delay.

In the top left panel of Figure **7.3** is the plot of the change in Absorbed Delay vs. increments of Carrier Delay at ATL. The plot also shows an increasing trend. However, it implies that the longer Carrier Delay gives rise to shorter Absorbed Delay since Absorbed Delay is negative by definition.

As can be seen from the above two figures, the changes in Airport Generated Delay and Airport Delay increase as the adjustment of Carrier Delay increases. The magnitude of the change is different for different airports.

Since the plots of Generated Delay variation in Figure **7.2** show that the relationship between delay variation and changes of Carrier Delay is monotonic, the univariate linear regression models can provide an approximate measure of the impact of changes in Carrier Delay on Generated Delay. The slope in these simple linear models represents the contribution of each unit of increment of Carrier Delay to Airport Generated Delay (in Table **7.1**). As can be seen from Table **7.1**, reducing 10.35 minute Carrier Delay, which is mean value of Carrier Delay, in June and July 2005 can reduce more than 4 minutes Generated Delay at the airports in the first column Table **7.1**. The reduction at remaining airports is less than 4 minutes.

Table **7.1**: Slopes of Changes in Airport Generated Delay vs. Percentage of Increment of Carrier Delay (mean 10.35 minutes) at 34 OEP Airports (minute). Airports are listed in the order of slopes.

Airport	**slope**	**Airport**	**slope**	**Airport**	**slope**	**Airport**	**slope**
ORD	5.77	LAX	3.41	IAH	2.48	TPA	1.47
ATL	4.71	SEA	3.21	LGA	2.38	SAN	1.42
MSP	4.34	CLT	2.88	CVG	2.26	PDX	1.33
PHL	4.30	LAS	2.85	MDW	2.23	MEM	1.33
PHX	4.29	IAD	2.75	SFO	2.10	CLE	1.06
DEN	4.16	BOS	2.66	DCA	2.04		
DTW	4.01	MIA	2.59	BWI	1.81		
JFK	3.91	MCO	2.53	SLC	1.74		
EWR	3.70			STL	1.74		
DFW	3.68			FLL	1.61		
				PIT	1.59		

The increase of Airport Generated Delay is consistently higher than the increase of Absorbed Delay. The plots in Figure **7.3** use smaller scale of y-axis than in Figure **7.2** in order to see the changes. Given such small amount of changes in Absorbed Delay, the slopes for the Absorbed Delay are not calculated.

7.2.2 GDP Holding Time

The value of GDP Holding Time is non-negative in the CATSR delay database. Only 23.1% of the epochs in the training set has greater than zero value of GDP Holding Time. Hence, the adjustment of GDP Holding Time is similar to Carrier Delay described in the previous section. The mean value of GDP Holding Time is the aggregate value of all positive GDP Holding Time, 19.22 minutes, as calculated in Equation **7.6**.

$$\text{Mean GDP Holding Time} = \frac{\sum_{airport=1}^{34}\sum_{day=1}^{61}\sum_{epoch=24}^{87} \text{GDP Holding Time}_{airport,day,epoch}}{\sum_{airport=1}^{34}\sum_{day=1}^{61}\sum_{epoch=24}^{87} I_{airport,day,epoch}}$$

$$= 19.22\text{(minute)} \qquad \textbf{7.6}$$

where, $\text{GDP Holding Time}_{\text{airport,day,epoch}}$ is average GDP Holding Time at an epoch in a day at an airport

$$I_{epoch} = \begin{cases} 1, \ \text{GDP Holding Time}_{\text{airport,day,epoch}} > 0 \\ 0, \ \text{otherwise} \end{cases}$$

The process of the adjustment on positive GDP Holding Time is formulated in Equation **7.7**. Only positive GDP Holding Time will be manipulated and the minimum value of GDP Holding Time is zero.

$$\text{adjusted GDP Holding Time}_{\text{airport,day,epoch}}$$
$$= \begin{cases} \max\left(\text{GDP Holding Time}_{\text{airport,day,epoch}} + w\% * \text{Mean GDP Holding Time}, 0\right), \\ \qquad \qquad \text{GDP Holding Time}_{\text{airport,day,epoch}} > 0 \\ 0, \ \text{otherwise} \end{cases} \qquad \textbf{7.7}$$

where, $w \in \{-50,-40,-30,-20,-10,10,20,30,40,50\}$

Delay Variation at 34 OEP airports

Figure **7.4** displays the changes in Airport Generated Delay as a result of changes in GDP Holding Time. No plot is shown in Figure **7.5** for the 5 airports in which GDP Holding Time is not a predictor.

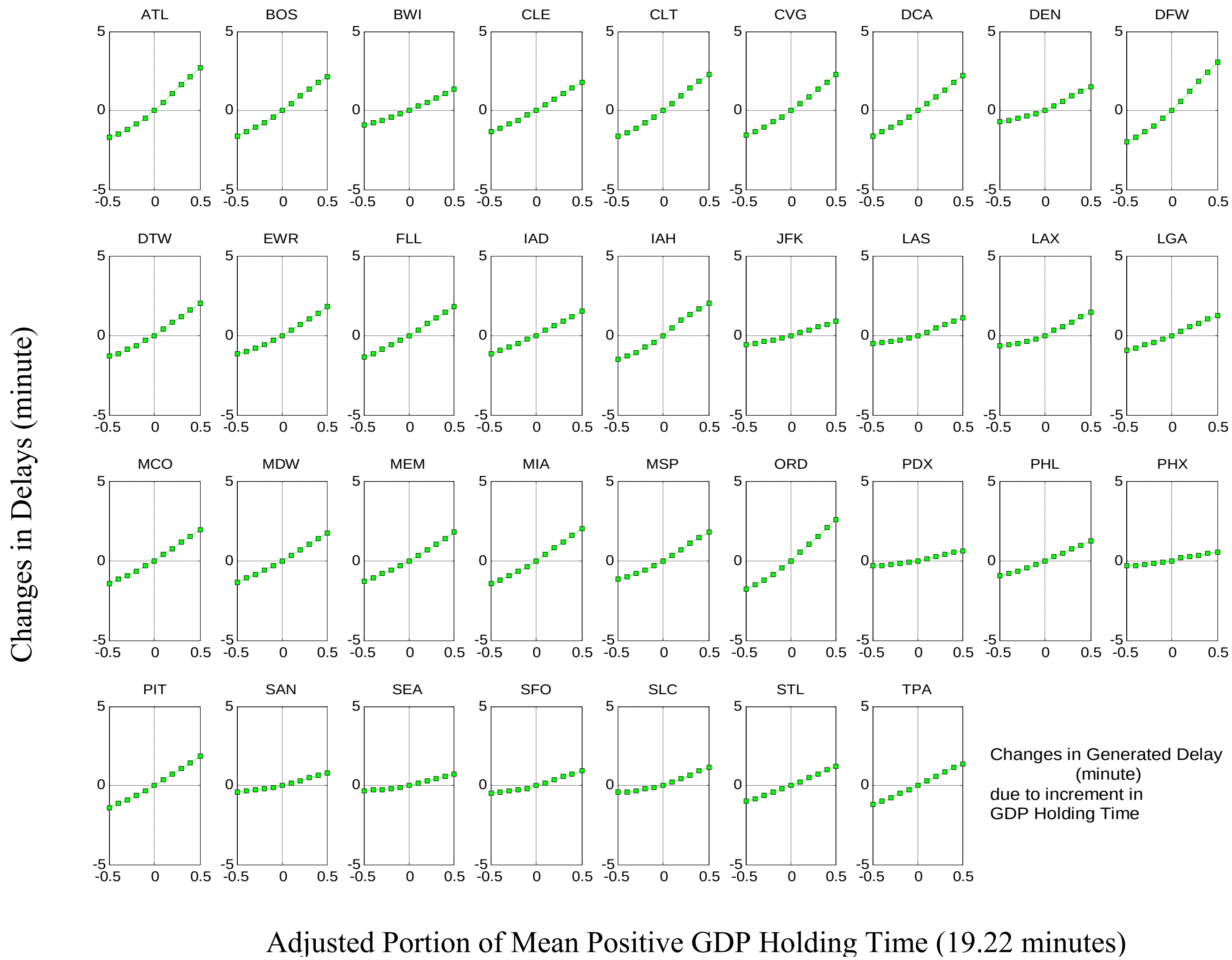


Figure **7.4**: Changes in Airport Generated Delay from Adjustment of GDP Holding Time (minute)

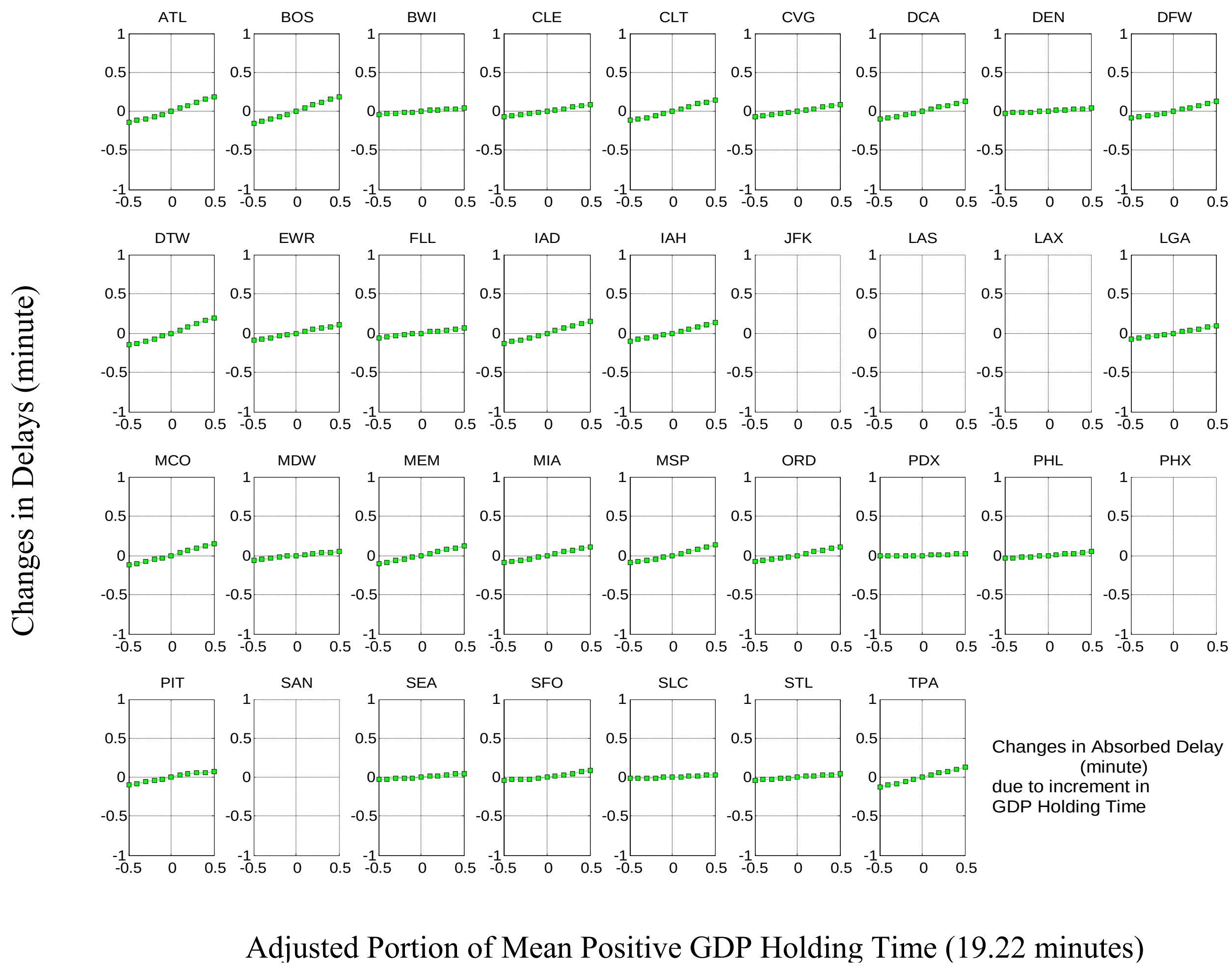


Figure **7.5**: Changes in Airport Absorbed Delay from Adjustment of GDP Holding Time (minute)

Similar to Carrier Delay, there is a clear upward trend of delay variation with greater adjustments in the GDP Holding Time. Increasing GDP Holding Time increases Airport Generated Delay and reduces the magnitude of Absorbed Delay. The slopes of the curves in Figure **7.4** are reported in Table **7.2**

Table **7.2**: Slopes of Airport Generated Delay Variation vs. Increment of GDP Holding Time (mean 19.22 minutes) at 34 OEP Airports (minute). Airports are listed in the order of slopes.

Airport	Slope	Airport	Slope	Airport	Slope	Airport	Slope
DFW	5.15	DTW	3.41	BWI	2.31	SFO	1.49
ATL	4.50	MCO	3.39	DEN	2.28	JFK	1.47
ORD	4.47	PIT	3.26	STL	2.26	SAN	1.20
CLT	4.01	FLL	3.25	PHL	2.24	SEA	1.11
DCA	3.92	CLE	3.15	LGA	2.21	PDX	0.98
CVG	3.88	MDW	3.13	LAX	2.20	PHX	0.95
BOS	3.87	MEM	3.07	LAS	1.68		
IAH	3.73	MSP	3.04	SLC	1.61		
MIA	3.51	EWR	2.98				
		IAD	2.69				
		TPA	2.63				

As for the changes in Generated Delay, the slopes for the airports in the last column of Table **7.2** are smaller than those for other airports. Adding 19.22 minutes of GDP Holding Time to these airports only increases overall Generated Delay by around 1 minute. With the exception of JFK, these airports are located at the west coast of the U.S. The changes in Absorbed Delay are so small that the scale in Figure **7.5** has to be reduced to -1 to 1. For this reason, the slopes of changes in Absorbed Delay are not calculated.

Table **7.3** shows the percentage of epochs in June and July 2005 when the GDP was issued at each airports and the mean values of these positive GDP Holding Times.

Table **7.3**: Percentage of GDP and Mean Value of GDP Holding Time at 34 OEP Airports in June and July 2005. Airports are listed in the order of percentage.

Airport	Percentage	Mean (minute)	Airport	Percentage	Mean (minute)
DFW	37.2%	14.3	TPA	22.9%	30.1
ORD	35.4%	15.5	LAX	22.9%	6.3
ATL	34.2%	14.1	LGA	22.3%	20.3
IAH	30.4%	15.7	CLE	22.3%	26.2
DTW	29.8%	19.4	PIT	22.2%	31.6
BOS	29.5%	20.5	MEM	20.9%	27.9
CLT	29.5%	24.5	MDW	19.4%	26.1
MCO	29.3%	22.9	BWI	17.7%	23.2
DCA	29.3%	20.8	PHX	17.5%	8.0
MSP	29.0%	15.6	LAS	17.5%	8.5
MIA	27.0%	28.4	STL	16.5%	29.7
EWR	25.6%	21.8	SFO	15.6%	9.1
FLL	25.5%	26.2	SEA	14.4%	11.0
PHL	25.3%	21.6	SLC	14.3%	11.9
CVG	24.7%	20.4	JFK	12.7%	16.7
IAD	24.6%	23.4	SAN	10.7%	14.6
DEN	23.5%	8.7	PDX	7.1%	20.9

As can be seen, the airports listed in the first group in Table **7.2** are listed near the top of Table **7.3**. These airports have high percentage of GDP. CLT has long GDP Holding Time (24.5 minutes), and the influence of GDP on its Generated Delay is ranked at 4th place.

Changing GDP Holding Time has very small impact on the Airport Absorbed Delay, so we do not discuss its influence in detail.

7.2.3 Ratio of Departure Demand and Capacity (Departure Demand Ratio)

The values of Departure Demand Ratio in Airport Generated Delay models is calculated by averaging the Departure Demand Ratio of all flights scheduled to depart at each 15-minute epoch of a day. Each flight's Departure Demand Ratio is obtained by dividing the number of scheduled departures (pushing back from gate) by the number of actual takeoffs in the 30-minute window of that flight's scheduled push back time.

Departure Demand Ratio greater than 1 indicates airport congestion, i.e., demand exceeds capacity. Departure Demand Ratio less than 0.5 indicates the slot assignments at the airport were inappropriate, i.e., the airport departed flights 2 times more than originally scheduled. The analysis from historical data of summer 2005 shows that the relationship between Departure Demand Ratio and Airport Generated Delay is not monotonic. The minimum Airport Generated Delay occurs within range 0.5 to 0.6 across 34 airports. Therefore, we used 0.55 instead of zero as the cutting point for the adjusted value for our sensitivity analysis.

Figure **7.1** shows that the distribution of Departure Demand Ratio has a long-right-tailed bell shape. There were 0.67% epochs among 34 OEP airports having greater than 5 Departure Demand Ratio. Such a Departure Demand Ratio is associated with situations when there were flights scheduled to depart but no actual takeoffs. Such a skewed distribution might suggest use of the median rather than the means as a measure of central tendency. The value for the median is 1.05. However, the mean value is still used as the base for sensitivity analysis since we use mean value as base for all other factors. The mean of Departure Demand Ratio was calculated using Equation **7.8**.

$$\text{Mean Departure } \rho = \frac{1}{n}\sum_{airport=1}^{34}\sum_{day=1}^{61}\sum_{epoch=24}^{87} \text{Departure } \rho_{airport,day,epoch} = 1.25 \tag{7.8}$$

where, Departure $\rho_{\text{airport,day,epoch}}$ is average Departure ρ at an epoch in a day at an airport

n is total number of records in the training sample of 34 airports

The adjustment to Departure Demand Ratio is similar to the adjustment to Inbound Delay. We add increments to all Departure Demand Ratio, but only subtract from Ratios that are above 0.55 and constrained the adjusted values to be no less than 0.55. This process is described in Equation **7.9**.

$$\text{adjusted Departure } \rho_{\text{airport,day,epoch}} = \begin{cases} \max\left(\begin{array}{l}\text{Departure } \rho_{\text{airport,day,epoch}} \\ \quad + w\% * \text{Mean Departure } \rho,\ 0.55\end{array}\right), & \text{Departure } \rho_{\text{airport,day,epoch}} \geq 0.55 \\ \max\left(\begin{array}{l}\text{Departure } \rho_{\text{airport,day,epoch}} \\ \quad + w\% * \text{Mean Departure } \rho,\ \text{Departure } \rho_{\text{airport,day,epoch}}\end{array}\right), & \text{otherwise} \end{cases} \tag{7.9}$$

where, $w \in \{-50,-40,-30,-20,-10,10,20,30,40,50\}$

Delay Variation at 34 OEP airports

Departure Demand Ratio is a significant factor at 33 Airport Generated Delay models. In 23 models, the Departure Demand Ratio in 30-minute window is significant. At MEM, the Departure Demand Ratio in 15-minute window is significant. The Departure Demand and Airport Departure Rate (ADR) in either 30-minute or 15-minute window is significant in the Generated Delay model of other 10 airports. To conduct a fair comparison, only Ratio of Departure Demand and Departure Throughput in 30-minute window is analyzed in this section. Hence, MEM and other 10 airports having Departure Demand and Airport Departure Rate (ADR) as predictor have no plot in Figure **7.6**.

Figure **7.6** shows the increasing trend of the changes in Generated Delay vs. increments of Departure Demand Ratio. As we would expect, higher Departure Demand Ratio implies more congested airport, and more congested airport has longer delays. Table **7.4** listed the slopes in the univariate regression models fitted to the dots in Figure **7.6**. The degree of influence of Departure Demand Ratio at airport PHL, LGA, EWR, and ATL is higher than other airports.

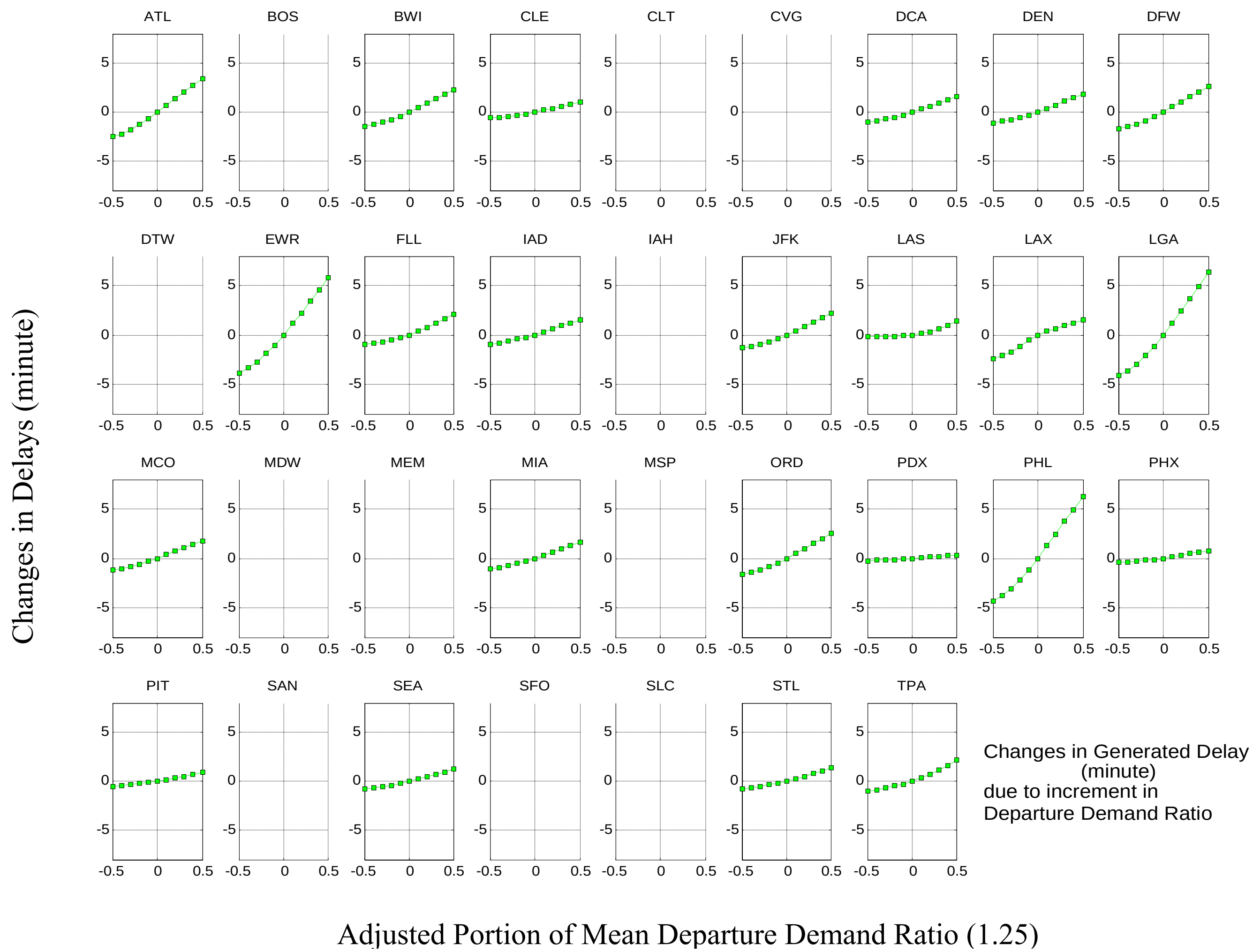


Figure **7.6**: Changes in Airport Generated Delay from Adjustment of Ratio of Departure Demand and Capacity_30min

Table **7.4**: Slopes of Changes in Airport Generated Delay vs. Increments of Departure Demand Ratio (mean 1.25) at 34 OEP Airports (minute). Airports are listed in the order of slopes.

Airport	Slope	Airport	Slope	Airport	Slope	Airport	Slope
PHL	10.94	DFW	4.50	TPA	3.05	STL	2.13
LGA	10.73	ORD	4.29	MCO	3.03	SEA	2.07
EWR	9.93	LAX	4.14	FLL	3.01	CLE	1.65
ATL	6.16	BWI	3.85	DEN	2.99	LAS	1.49
		JFK	3.58	MIA	2.82	PIT	1.42
				DCA	2.73	PHX	1.25
				IAD	2.55	PDX	0.57

7.2.4 Airline Swap Aircraft Rate

In Airport Delay database, only 4.1% epochs of 34 OEP airports has greater than zero Swap Aircraft Rate value. Swap Aircraft Rate is a significant factor at 10 Airport Generated Delay models. Similar to Carrier Delay and GDP Holding Time, the mean value of Swap Aircraft Rate calculated for sensitivity analysis is the aggregated value of positive Swap Aircraft Rate in Equation **7.10**. The adjustment process of positive Swap Aircraft Rate is formulated in Equation **7.11**.

$$\text{Mean SwapAircraftRate} = \frac{\sum_{airport=1}^{34} \sum_{day=1}^{61} \sum_{epoch=24}^{87} \text{SwapAircraftRate}_{airport,day,epoch}}{\sum_{airport=1}^{34} \sum_{day=1}^{61} \sum_{epoch=24}^{87} I_{airport,day,epoch}}$$

$$= 0.23 \qquad \textbf{7.10}$$

where, $\text{SwapAircraftRate}_{\text{airport,day,epoch}}$ is average SwapAircraftRate at an epoch in a day at an airport

$$I_{epoch} = \begin{cases} 1, & \text{SwapAircraftRate}_{\text{airport,day,epoch}} > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$\text{adjusted SwapAircraftRate}_{\text{airport,day,epoch}}$$

$$= \begin{cases} \max\left(\text{SwapAircraftRate}_{\text{airport,day,epoch}} + w\% * \text{Mean SwapAircraftRate},\ 0\right), & \text{SwapAircraftRate}_{\text{airport,day,epoch}} > 0 \\ 0, & \text{otherwise} \end{cases} \qquad \textbf{7.11}$$

where, $\text{w} \in \{-50,-40,-30,-20,-10,10,20,30,40,50\}$

Delay Variation at 34 OEP airports

The scale of the y-axis has to be shrunk in Figure **7.7** since the variation is small. This is probably due to the small proportion of Swap Aircraft Rate in the whole data sample. Although the variation is very small, the plots still show some difference among airports. The increasing trend of ATL, DEN and ORD is steeper than other airports.

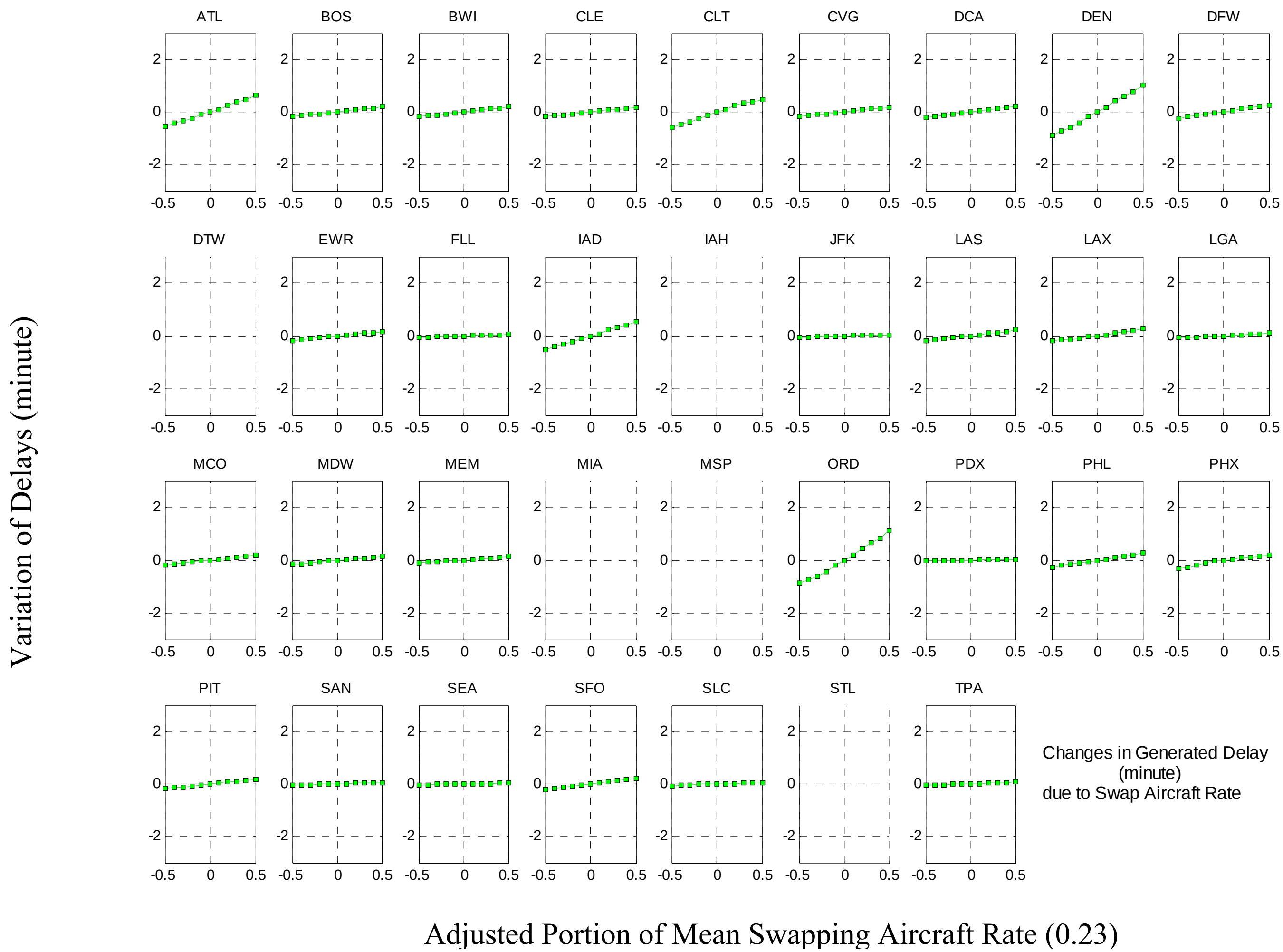


Figure **7.7**: Changes in Airport Generated Delay from Adjustment of Swap Aircraft Rate

7.2.5 Inbound Delay

There are negative Inbound Delays and positive Inbound Delays in the database. Negative Inbound Delay means the flights that arrived earlier than their scheduled time. From an operation perspective, the ideal value for Inbound Delay is zero, i.e., flights arrive on time. In summer 2005, only 16% epochs across 34 OEP airports had zero Inbound Delay. 44.9% epochs had positive Inbound Delay and 39.1% has negative Inbound Delay. Neither late arrival nor early arrival can use the originally assigned slots. Both of them should be avoided. Practically, late arrival will cause more problems than early arrival since it will disturb the connectivity of airline's schedule of crew and airframe.

The mean value of Inbound Delay is 20.65 minutes calculated by aggregating the values of positive Inbound Delay (in Equation **7.12**).

$$\text{Mean Inbound Delay} = \frac{\sum_{airport=1}^{34}\sum_{day=1}^{61}\sum_{epoch=24}^{87}\text{Inbound Delay}_{airport,day,epoch}}{\sum_{airport=1}^{34}\sum_{day=1}^{61}\sum_{epoch=24}^{87} I_{airport,day,epoch}} = 20.65(\text{minute}) \qquad \mathbf{7.12}$$

where, $\text{Inbound Delay}_{\text{airport,day,epoch}}$ is average Inbound Delay at an epoch in a day at an airport

$$I_{epoch} = \begin{cases} 1, & \text{Inbound Delay}_{\text{airport,day,epoch}} > 0 \\ 0, & \text{otherwise} \end{cases}$$

The process of adjusting Inbound Delay is formulated in Equation **7.13**. When we added a positive amount to Inbound Delay, it was added to all epochs. However, when we do subtraction, only positive Inbound Delay will be changed and we do not allow it to go below zero.

$$\text{adjusted Inbound Delay}_{\text{airport,day,epoch}} = \begin{cases} \max\left(\begin{aligned}&\text{Inbound Delay}_{\text{airport,day,epoch}} \\ &\quad + w\% * \text{Mean Inbound Delay},\ 0\end{aligned}\right), \\ \qquad\qquad\qquad\qquad \text{Inbound Delay}_{\text{airport,day,epoch}} \geq 0 \\ \max\left(\begin{aligned}&\text{Inbound Delay}_{\text{airport,day,epoch}} \\ &\quad + w\% * \text{Mean Inbound Delay},\ \text{Inbound Delay}_{\text{airport,day,epoch}}\end{aligned}\right), \\ \qquad\qquad\qquad\qquad \text{otherwise} \end{cases} \tag{7.13}$$

$$\text{where, } w \in \{-50,-40,-30,-20,-10,0,10,20,30,40,50\}$$

Delay Variation at 34 OEP airports

In Figure **7.8**, the curves of Generated Delay variation are either flat or have a downward trend as Inbound Delay increases, except at LGA. The curve is upward at LGA. This downward trend implies that the Inbound Delay results in shorter Generated Delay at the airport. That may due to an airline’s recovery strategy to prevent delay getting longer. This phenomenon is illustrated more clearly in Figure **7.9**. The plots show that the magnitude of Absorbed Delay becomes longer when Inbound Delay increases.

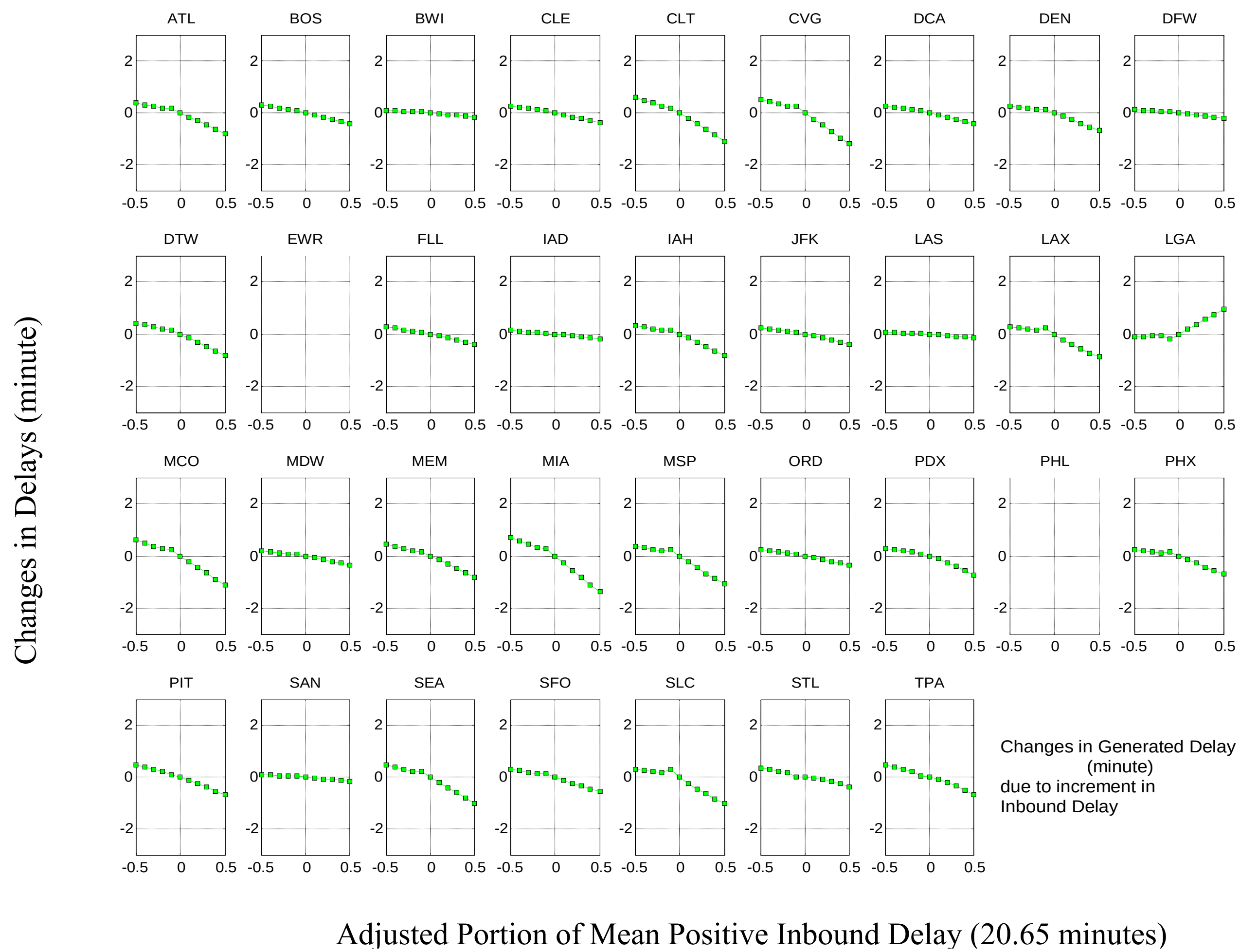


Figure **7.8**: Changes in Airport Generated Delay from Adjustment of Inbound Delay (minute)

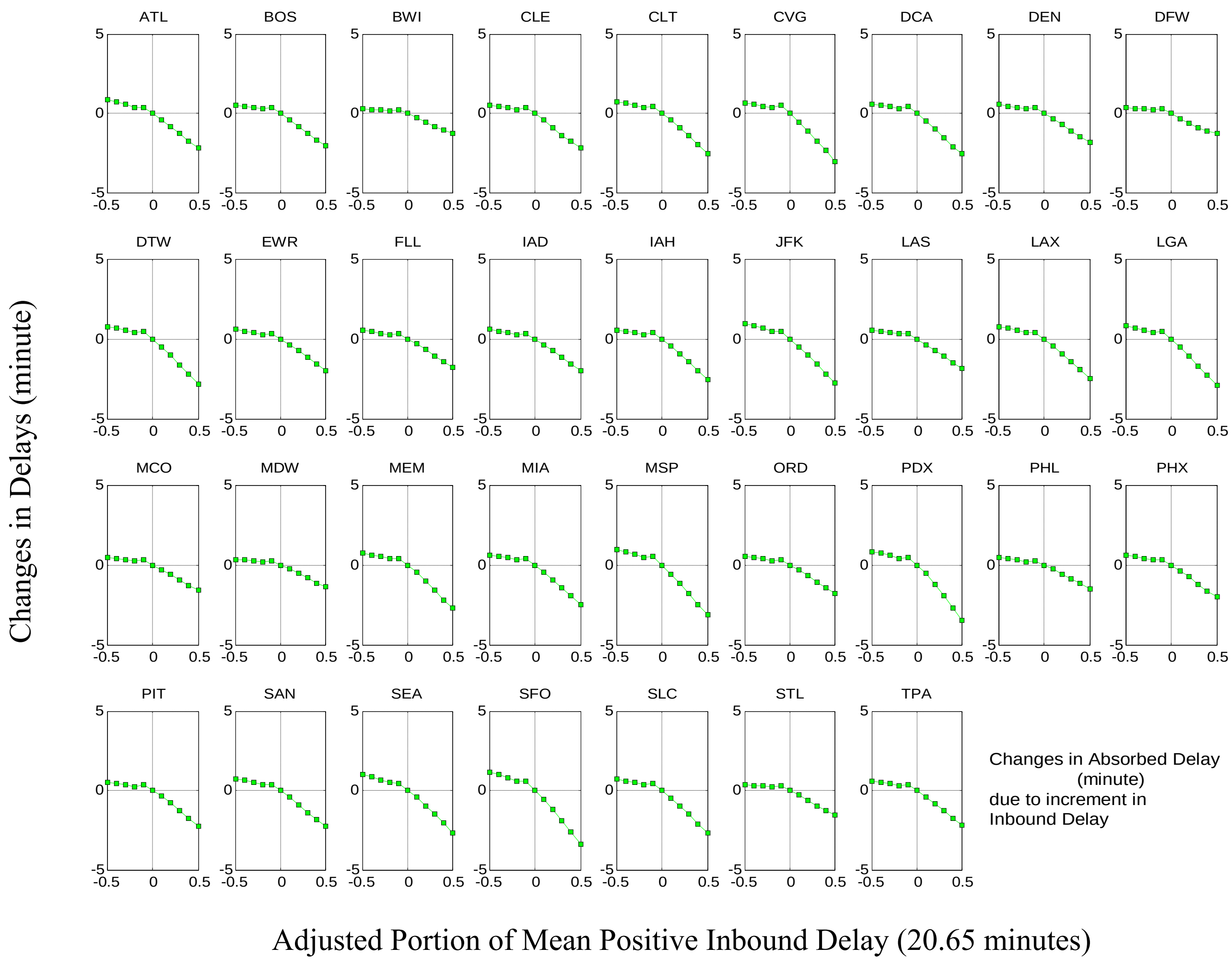


Figure **7.9**: Changes in Airport Absorbed Delay from Adjustment of Inbound Delay (minute)

The slopes of the univariate regression models fitted for the dots in each plot in Figure **7.9** ranges from -4.57 minutes to -1.65 minutes (in Table **7.5**).

Table **7.5**: Slopes of Airport Absorbed Delay Variation vs. Increments of Inbound Delay (mean 20.65) at 34 OEP Airports (minute). Airports are listed in the order of slopes.

Airport	**Slope**	**Airport**	**Slope**	**Airport**	**Slope**
SFO	-4.57	MEM	-3.48	LAS	-2.49
PDX	-4.28	SLC	-3.39	DEN	-2.47
MSP	-4.16	LAX	-3.30	ORD	-2.40
LGA	-3.77	DCA	-3.25	FLL	-2.37
JFK	-3.75	CLT	-3.25	MCO	-2.06
CVG	-3.69	MIA	-3.10	STL	-1.98
SEA	-3.66	ATL	-3.06	PHL	-1.95
DTW	-3.61	IAH	-3.06	DFW	-1.79
		SAN	-3.05	MDW	-1.77
		TPA	-2.82	BWI	-1.65
		CLE	-2.78		
		PIT	-2.72		
		BOS	-2.69		
		PHX	-2.67		
		EWR	-2.64		
		IAD	-2.60		

Given every 20.65-minute Inbound Delay, these airports can absorb it from 8.0% (1.65/20.65=8%) to 22.1% (4.57/20.65=22.1%). The mean absorbed percentage is 14.3% of 34 OEP airports, and the percentage of delay which is not absorbed is 85.7%. If there were no generated delay at the airport, this number is close to the analytical results from the research of Boswell and Evans (1997). They estimated that about 80% of delays propagate to downstream airports.

When a flight arrives late, it is common for airlines to make up schedule by reducing the turn-around time. The Scheduled Turn-around Time is investigated in the next section.

7.2.6 Scheduled Turn-around Time

Scheduled Turn-around Time is a significant factor for Airport Absorbed Delay model at all 34 OEP airports but not so at Airport Generated Delay models. The only constrain for the adjustment is that no turn-around time is allowed to be less than 0.

The mean value of Scheduled Turn-around Time is calculated using Equation **7.14** and the adjustment process is formulated in Equation **7.15**.

$$\begin{aligned}&\text{Mean Scheduled Turnaround Time}\\&\quad=\frac{1}{n}\sum_{airport=1}^{34}\sum_{day=1}^{61}\sum_{epoch=24}^{87}\text{Scheduled Turnaround Time}_{airport,day,epoch}\\&\quad=60.04(\text{minute})\end{aligned} \qquad \mathbf{7.14}$$

where, $\text{Scheduled Turnaround Time}_{\text{airport,day,epoch}}$ is average Scheduled Turnaround Time at an epoch in a day at an airport

n is the total number of records in June and July at 34 OEP airports

$$\begin{aligned}&\text{adjusted Scheduled Turnaround Time}_{\text{airport,day,epoch}}\\&=\max\left(\begin{array}{l}\text{Scheduled Turnaround Time}_{\text{airport,day,epoch}}\\ \qquad\qquad +w\%*\text{Mean Scheduled Turnaround Time},\ 0\end{array}\right)\end{aligned} \qquad \mathbf{7.15}$$

where, $w\in\{-50,-40,-30,-20,-10,10,20,30,40,50\}$

Delay Variation at 34 OEP airports

The plots in Figure **7.10** reveal that the absolute value of Absorbed Delay increases as turn-around prolonged. However, after adding 40% of turn-around time the increase of Absorbed Delay stops at DCA, DFW, ORD, and TPA.

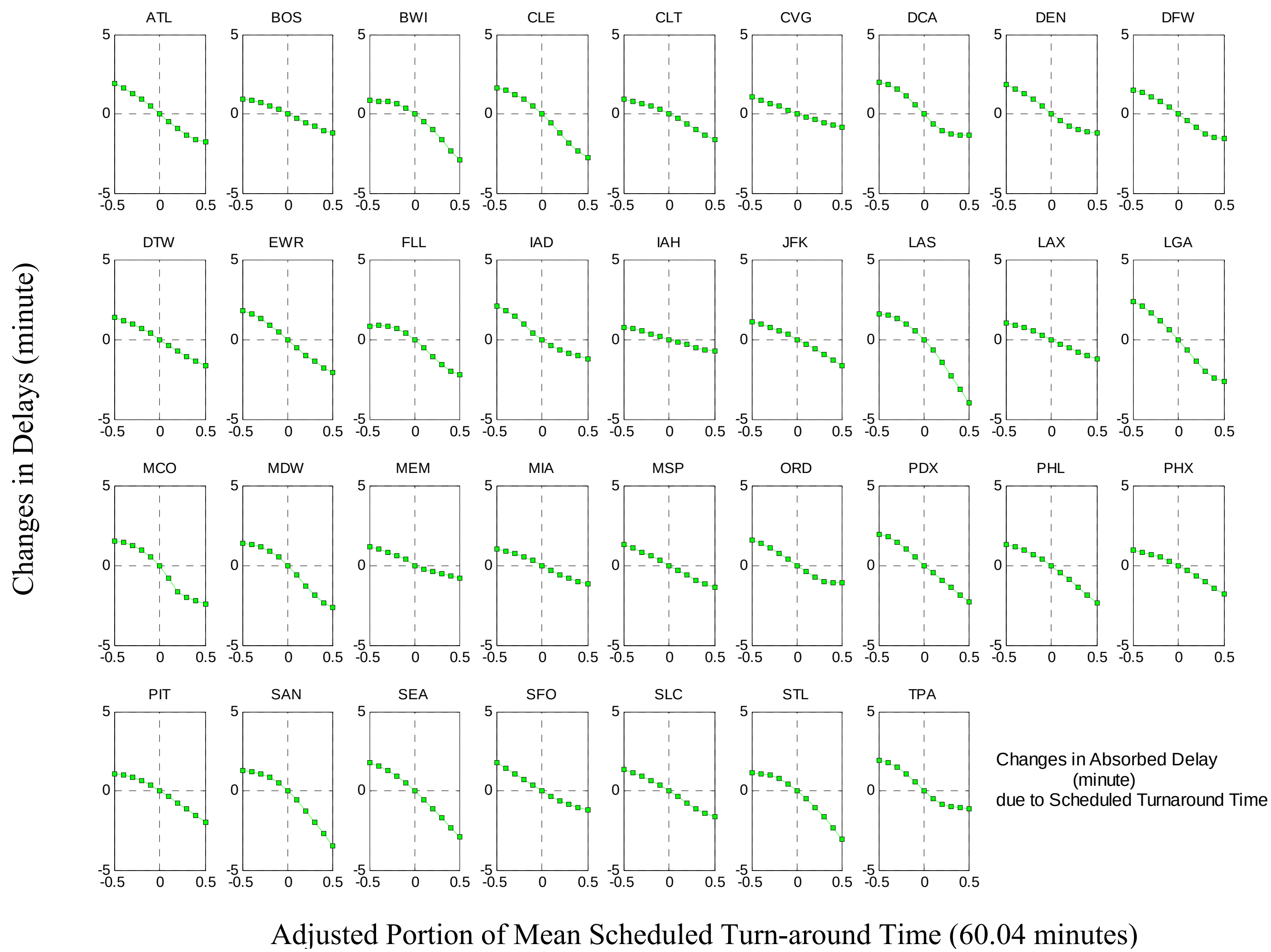


Figure **7.10**: Changes in Airport Absorbed Delay from Adjustment of Scheduled Turn-around Time

The slopes of the univariatie regression models fitted for the dots in each plot in Figure **7.10** ranges from -5.72 minutes to -1.60 minutes (in Table **7.6**). The negative value of slope means that increase turn-around time at the airports in the left column will increase the length of absorbed delays.

Table **7.6**: Slopes of Airport Absorbed Delay Variation vs. Increments of Scheduled Turnaround Time (mean 60.04 minutes) at 34 OEP Airports (minute). Airports are listed in the order of slopes.

Airport	**Slope**	**Airport**	**Slope**	**Airport**	**Slope**
LAS	-5.72	FLL	-3.46	CVG	-1.95
LGA	-5.47	DFW	-3.42	IAH	-1.60
SAN	-4.88	DEN	-3.37		
SEA	-4.82	DTW	-3.23		
CLE	-4.70	PIT	-3.21		
MCO	-4.62	SLC	-3.17		
MDW	-4.47	SFO	-3.11		
PDX	-4.47	ORD	-3.05		
STL	-4.32	JFK	-2.83		
EWR	-4.12	PHX	-2.79		
ATL	-4.04	MSP	-2.79		
DCA	-3.95	CLT	-2.66		
BWI	-3.87	LAX	-2.42		
PHL	-3.79	MIA	-2.36		
TPA	-3.53	BOS	-2.32		
IAD	-3.51	MEM	-2.11		

7.2.7 Number of Seats

Number of Seats is a significant factor for Airport Absorbed Delay model at all 34 OEP airports but not so at Airport Generated Delay models. Its minimum value in the data of summer of 2005 is 30. Hence, the constraint for the adjustment is that the value of Number of Seats after adjustment is not allowed to be less than 30.

The mean value of Number of Seats is calculated using Equation **7.16** and the adjustment process is formulated in Equation **7.17**.

$$\begin{aligned}&\text{Mean Number of Seats}\\&\quad=\frac{1}{n}\sum_{airport=1}^{34}\sum_{day=1}^{61}\sum_{epoch=24}^{87}\text{Number of Seats}_{airport,day,epoch}\\&\quad=134.43\end{aligned} \qquad \mathbf{7.16}$$

where, $\text{Number of Seats}_{\text{airport,day,epoch}}$ is average Number of Seats at an epoch in a day at an airport

n is the total number of records in 2 months at 34 OEP airports

$$\begin{aligned}&\text{adjusted Number of Seats}_{\text{airport,day,epoch}}\\&=\max\left(\text{Number of Seats}_{\text{airport,day,epoch}}+w\%*\text{Number of Seats},\ 30\right)\end{aligned} \qquad \mathbf{7.17}$$

where, $w\in\{\text{-50,-40,-30,-20,-10,10,20,30,40,50}\}$

Delay Variation at 34 OEP airports

The plots in Figure **7.11** reveal that the absolute value of Absorbed Delay increases as the Number of Seats increased. The positive slope means that the length of Absorbed Delay reduces as the number of seats increases since the value of Absorbed Delay is negative. In other words, more delay is absorbed when the number of seats on the aircraft is smaller. This is probably because the smaller aircraft is able to turn around faster.

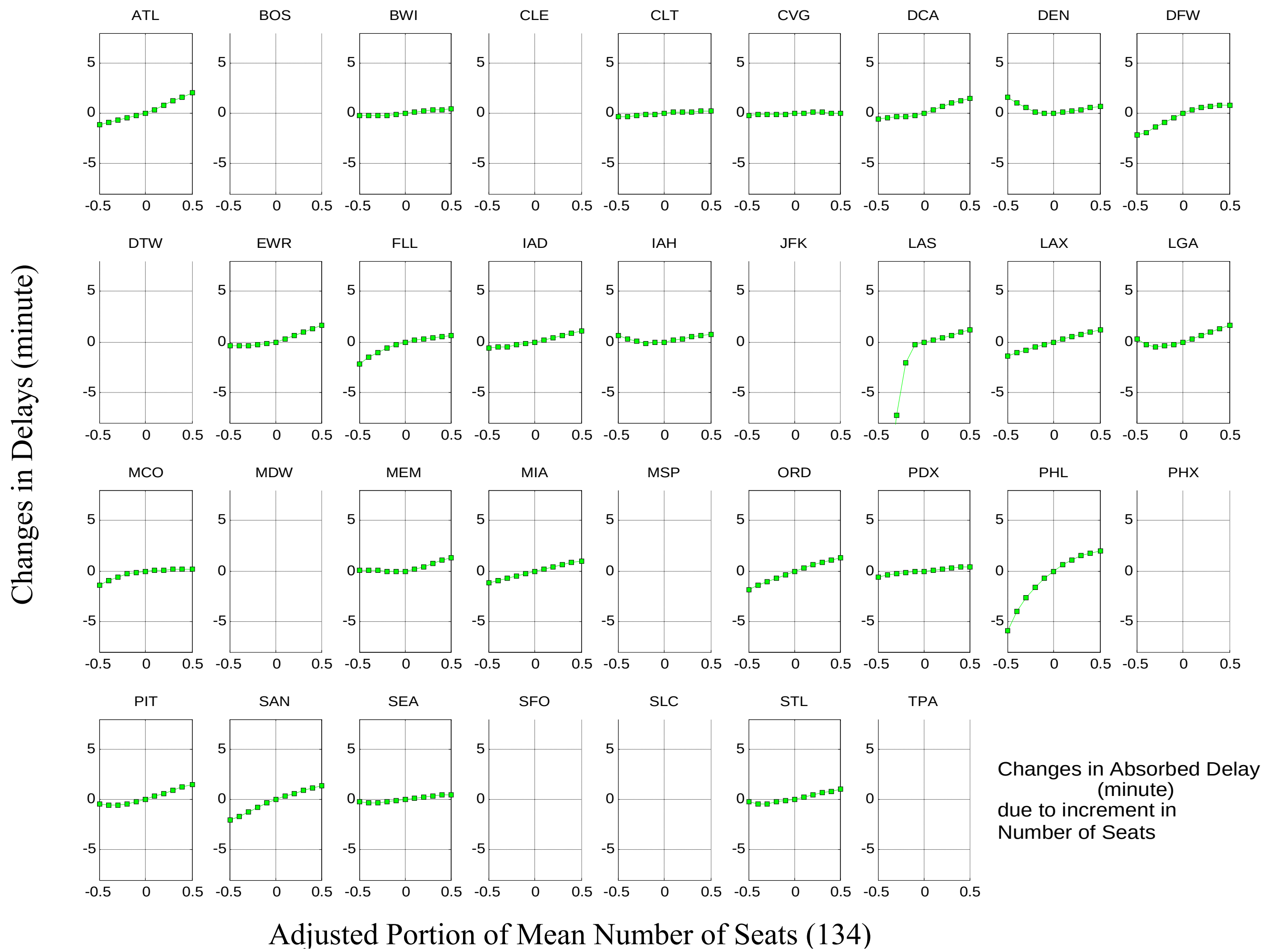


Figure **7.11**: Changes in Airport Absorbed Delay from Adjustment of Number of Seats (mean=134)

The curve for airport DEN, IAH, LGA and STL are not monotonic. Adding or removing 67 seats (0.5 of mean 134) will both reduce the length of the Absorbed Delay. The plot for airport LAS is quiet different from the others. A detailed investigation found that the second minimum value of seats at LAS is 100 and mean is 150. There is only one case with the average seats value of 70 at LAS. Removing more than 40% of average number of seats (54) from the data of LAS causes a large number of cases outside the value range of Number of Seats, i.e. less than 100. Hence, the left two points in the plot of LAS was left out. The slopes of the plots in Figure **7.11** are given in Table **7.7**.

Table **7.7**: Slopes of Airport Absorbed Delay Variation vs. Increments of Number of Seats (mean 134) at 34 OEP Airports (minute). Airports are listed in the order of slopes.

Airport	Slope	Airport	Slope	Airport	Slope
LAS	7.51	SAN	3.41	MCO	1.42
PHL	7.44	ATL	3.21	MEM	1.18
		DFW	3.21	PDX	1.00
		ORD	3.19	SEA	0.88
		FLL	2.67	BWI	0.81
		LAX	2.57	CLT	0.62
		MIA	2.21	IAH	0.43
		PIT	2.21		
		DCA	2.19		
		EWR	2.15		
		LGA	1.89		
		IAD	1.71		
		STL	1.51		

7.3 Comparison of Individual Factor's impact

7.3.1 Comparison of Factors of Airport Generated Delay

Most plots of Airport Generated Delay Variation demonstrate a monotonic relationship between individual factors and Airport Generated Delay. The slopes of fitted univariate regression models for the dots in the Delay Variation vs. Factor's Increments figures approximately represent the degree of influence from one factor. Table **7.8** shows the slopes calculated in the previous subsections on the factors influencing Airport Generated Delay.

Table **7.8**: Summary of Slopes of the Changes of Airport Generated Delay vs. Increments of factors at 34 OEP Airports. Airports are listed in the order of average airport delay in summer 2005.

	Average Airport Delay (min)	Inbound Delay	Departure Demand Ratio	Carrier Delay	Swap Aircraft Rate	GDP Time
PHL	25.14		10.94	4.30	0.52	2.24
JFK	22.56	-0.63	3.58	3.91	0.09	1.47
EWR	20.67		9.93	3.70	0.33	2.98
ORD	17.99	-0.61	4.29	5.77	2.02	4.47
MSP	16.80	-1.51		4.34		3.04
MIA	15.89	-2.13	2.82	2.59		3.51
IAD	15.83	-0.34	2.55	2.75	1.02	2.69
IAH	15.62	-1.17		2.48		3.73
LGA	14.65	1.07	10.73	2.38	0.16	2.21
DTW	14.32	-1.26		4.01		3.41
CLT	14.19	-1.68		2.88	1.12	4.01
ATL	13.80	-1.19	6.16	4.71	1.19	4.50
DFW	13.26	-0.32	4.50	3.68	0.49	5.15
BOS	12.33	-0.74		2.66	0.36	3.87
PHX	11.45	-0.96	1.25	4.29	0.52	0.95
FLL	11.41	-0.65	3.01	1.61	0.10	3.25
DCA	11.09	-0.70	2.73	2.04	0.42	3.92
MDW	11.05	-0.55		2.23	0.30	3.13
CVG	10.14	-1.74		2.26	0.36	3.88
MEM	9.96	-1.29		1.33	0.22	3.07
BWI	9.68	-0.27	3.85	1.81	0.37	2.31
LAS	9.20	-0.21	1.49	2.85	0.39	1.68
CLE	9.08	-0.66	1.65	1.06	0.35	3.15
TPA	8.44	-1.13	3.05	1.47	0.12	2.63
DEN	8.39	-0.96	2.99	4.16	1.93	2.28
PIT	8.32	-1.15	1.42	1.59	0.36	3.26
MCO	8.30	-1.73	3.03	2.53	0.36	3.39
SEA	8.17	-1.49	2.07	3.21	0.07	1.11
SLC	7.76	-1.41		1.74	0.09	1.61
STL	6.87	-0.67	2.13	1.74		2.26
LAX	6.74	-1.21	4.14	3.41	0.46	2.20
SFO	6.41	-0.88		2.10	0.42	1.49
SAN	5.35	-0.26		1.42	0.10	1.20
PDX	4.21	-1.01	0.57	1.33	0.07	0.98

These factors can be grouped into categories based on their slope value as shown in Figure **7.12**. The x-axes are the range of the slopes shown in Table **7.8** and the y-axes are number of airports associated with the corresponding slopes in x-axes. We define factor with slope greater than 5.5 as very high influence factor, slope 3.5 to 5.5 as high influence factor, slope 1.5 to 3.5 as moderate and slope less than 1.5 as low influence factor.

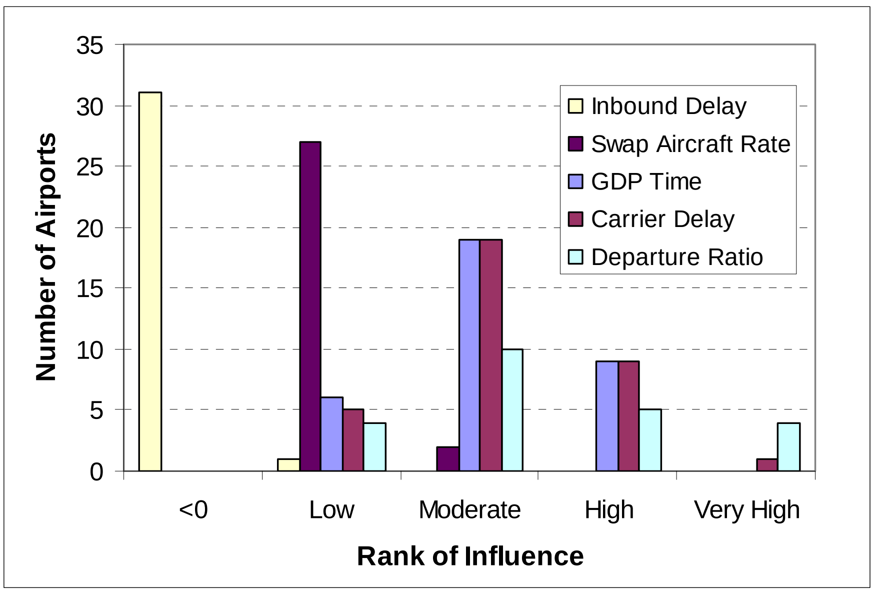


Figure **7.12**: Influence Rank of Factor of Generated Delay Based on the Slopes of the Changes of Generated Delay

Four airports are very highly influenced by Departure Demand Ratio; one of them is affected by Carrier Delay also. Nine airports are under high influence of Carrier Delay, and or GDP Holding Time, and 5 airports are under high influence of Departure Demand Ratio. Inbound Delay and Swap Aircraft Rate have low influence on the Generated Delay.

7.3.2 Comparison of Factors of Airport Absorbed Delay

For Airport Absorbed Delay, a longer value means stronger capability to make up delay. Slopes of Airport Absorbed Delay variation were calculated in Table **7.9**. The Absorbed Delay is defined as negative delay; the slope of beneficial factor is negative.

Table **7.9**: Summary of Slopes of the Changes of Airport Absorbed Delay vs. Increments of factors at 34 OEP Airports. Airports are listed in the order of average airport delay in summer 2005.

	Average Airport Delay (min)	Inbound Delay	Turnaround Time	Carrier Delay	Number of Seats	GDP Time
PHL	25.14	-1.95	-3.79	0.66	7.44	0.08
JFK	22.56	-3.75	-2.83	0.80		
EWR	20.67	-2.64	-4.12	0.49	2.15	0.20
ORD	17.99	-2.40	-3.05	0.61	3.19	0.20
MSP	16.80	-4.16	-2.79	0.45		0.23
MIA	15.89	-3.10	-2.36	0.29	2.21	0.21
IAD	15.83	-2.60	-3.51	0.25	1.71	0.29
IAH	15.62	-3.06	-1.60	0.24	0.43	0.24
LGA	14.65	-3.77	-5.47	0.36	1.89	0.17
DTW	14.32	-3.61	-3.23	0.52		0.35
CLT	14.19	-3.25	-2.66	0.62	0.62	0.28
ATL	13.80	-3.06	-4.04	0.49	3.21	0.34
DFW	13.26	-1.79	-3.42	0.21	3.21	0.21
BOS	12.33	-2.69	-2.32	0.23		0.36
PHX	11.45	-2.67	-2.79	0.47		
FLL	11.41	-2.37	-3.46	0.43	2.67	0.13
DCA	11.09	-3.25	-3.95	0.45	2.19	0.23
MDW	11.05	-1.77	-4.47	0.28		0.11
CVG	10.14	-3.69	-1.95	0.20	0.25	0.16
MEM	9.96	-3.48	-2.11	0.23	1.18	0.24
BWI	9.68	-1.65	-3.87	0.20	0.81	0.08
LAS	9.20	-2.49	-5.72	0.39	7.51	
CLE	9.08	-2.78	-4.70	0.18		0.15
TPA	8.44	-2.82	-3.53	0.20		0.26
DEN	8.39	-2.47	-3.37	0.54	-0.60	0.06
PIT	8.32	-2.72	-3.21	0.14	2.21	0.18
MCO	8.30	-2.06	-4.62	0.23	1.42	0.27
SEA	8.17	-3.66	-4.82	0.48	0.88	0.07
SLC	7.76	-3.39	-3.17	0.18		0.04
STL	6.87	-1.98	-4.32	0.12	1.51	0.08
LAX	6.74	-3.30	-2.42	0.52	2.57	
SFO	6.41	-4.57	-3.11	0.72		0.12
SAN	5.35	-3.05	-4.88	0.41	3.41	
PDX	4.21	-4.28	-4.47	0.36	1.00	0.03

The value of slopes for factors of Absorbed Delay are separated using the same cutting point as we did for Generated Delay. An absolute value greater than 5.5 is defined as a very high influence factor, 3.5 to 5.5 as high influence factor, 1.5 to 3.5 as moderate and slope less than 1.5 as low influence factor. As can be seen from Figure **7.13**, both Carrier Delay and GDP Holding Time have very small influence on Absorbed Delay. Sixteen airports are highly impacted by the Scheduled Turn-around Time, and 8 airports are highly influenced by the Inbound Delay. The Number of Seats only have high influence on 2 airports.

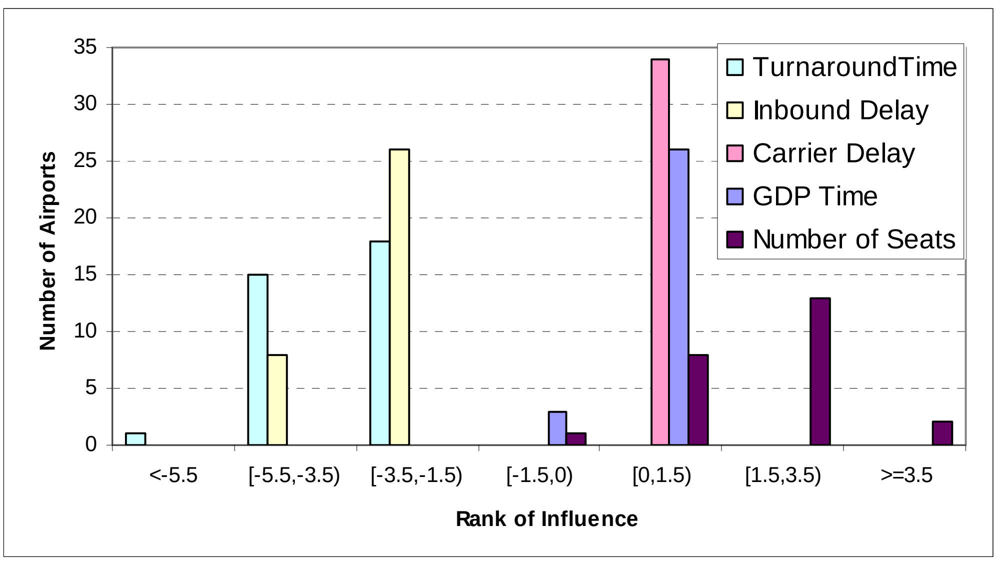


Figure **7.13**: Influence Rank of Factor of Absorbed Delay Based on the Slopes of the Changes of Absorbed Delay

7.4 Caveats

The sensitivity analysis conducted in this research has some limitations and modifications on standard sensitivity analysis. The limitations are:

1) For GDP Holding Time, Carrier Delay, and Swap Aircraft Rate, the adjustments are done on the magnitudes of these variables, not whether there is non-zero GDP, Carrier Delay, or Swapping Aircraft (greater than zero). So, not all cases in the data sets were adjusted.
2) The expected value of GDP Holding Time, Carrier Delay, Swap Aircraft Rate, and Inbound Delay are calculated using the mean of greater than zero values.
3) The special constraints were applied on the adjustment of each factor to force the data within the reasonable range of values. For example, the adjusted value for GDP Holding Time, Carrier Delay, Swap Aircraft Rate, Inbound Delay, and Scheduled Turn-around Time was cut off at zero. The Departure Demand Ratio is cut off at 0.55 and the Number of Seats is cut off at 30.
4) Regression models provide information about statistical relationships in data. Using the results of the sensitivity analysis to estimate the changes that would occur if the distributions of the input variables changed requires the assumption that the model reflects causal relationships. Statistically valid estimation of causal influences would require data from a study in which variables are manipulated. In the absence of experimental data, there results must be treated with caution.

The results of sensitivity analysis can only be used as a type of quantitative reference for policy makers. More information should be considered when making a policy change.

CHAPTER 8

CASE STUDY

This chapter describes the utilization of the delay models. These models were trained with historical data from June and July 2005. The form of the models and the process for developing the models are explained in chapter 3 and chapter 4. In this chapter, the final models reported in Appendix B were used in a case study of delay reduction policies.

8.1 Design of the Case Study

The case study was conducted by following 5 steps.

1. Select airports.

Seven airports (ATL, DEN, EWR, JFK, ORD, PHL, and LGA) were selected to represent characteristics of different airport classes. The arrival delays of outbound flights from these airports collectively account for 37% of total arrival delay at outbound destination from 34 OEP airports.

In the selected 7 airports, 6 of them (LGA is an exception) are among the 30 busiest airports in the world (Odoni 2004). PHL, JFK, EWR, and ORD have the highest Generated Delay per flight among the 34 OEP airports in 2005 (see Table **A.2**). ATL is the airport with the highest number of operations (see Table **A.1**). LGA has the highest Generated Delay among the non-hub airports. DEN is the largest international airport in the United States, whose land area is about 10 times of the land area of ATL (Odoni 2004). These airports also represent a range of geographical locations in the U.S.

2. Compare the impact of 3 significant control factors.

The factors analyzed are (1) GDP Holding Time, (2) Carrier Delay, and (3) Departure Demand Ratio. The settings for each factor are set below the mean of positive values of these factors.

- Set GDP Holding Time to 0, 4, 8 and 15 minutes. The mean GDP Holding Time is 19.22 minutes.
- Set Carrier Delay to 0, 5 and 10 minutes. The mean Carrier Delay is 10.35 minutes.
- Set Departure Demand Ratio to 0.8, 0.9, 1.0 and 1.2. The mean Departure Demand Ratio is 1.25.

3. Set departure time.

Heavy delay time (6pm to 6:15pm; epoch 72), was set as the scheduled departure time.

4. Set mean values for the model inputs.

The mean values of other inputs were calculated using the data from these 7 airports in June and July 2005. The mean value of each airport was set as the fixed value in the models. The purpose of calculating the sample mean for these input variables is to set them at typical values. The results from this case study are only valid within the normal range of these factors. Table **8.1** shows the set value for the inputs in the Generated Delay model of each airport.

Table **8.1**: Settings for Predictors at Generated Delay Model at each Airport

	PHL	JFK	EWR	ORD	LGA	ATL	DEN
	27.2	22.6	20.7	20.3	14.6	13.8	8.4
Inbound Delay (min)		8.62		18.96	27.19	27.67	8.81
Terminal Weather				0	0		
Swap Aircraft Rate	0.0241	0.0197	0.0095	0.0249	0.0112	0.0144	0.0068
Actual Enroute Time Weather	3794.44			857.68	859.67		
AAR							29.0
Schedule Enroute Time Weather	4246.79						
Visibility							9.758
Runway Configuration	2			1			
Arrival Demand Ratio (throughput 30min)	0.99						
Arrival Demand Ratio (AAR 15min)							0.43

Note: the value for Runway Configuration is the normalized value based on the mean taxi-out delay. 1 represents the lowest group of taxi-out delay and 3 represents the highest group of taxi-out delay.

5. Calculate and compare the predictions of Generated Delay for these scenarios.

Table **8.2** lists the scenarios studied in this chapter.

Table **8.2**: Scenarios

Control Variables			Prediction of Generated Delay
GDP=0 minute	Carrier Delay=0 minute	rho=0.8	
		rho=0.9	
		rho=1.0	
		rho=1.2	
	Carrier Delay=5 minutes	rho=0.8	
		rho=0.9	
		rho=1.0	
		rho=1.2	
	Carrier Delay=10 minutes	rho=0.8	
		rho=0.9	
		rho=1.0	
		rho=1.2	
GDP=4 minutes	Carrier Delay=0 minute	rho=0.8	
		rho=0.9	
		rho=1.0	
		rho=1.2	
	Carrier Delay=5 minute	rho=0.8	
		rho=0.9	
		rho=1.0	
		rho=1.2	
	Carrier Delay=10 minutes	rho=0.8	
		rho=0.9	
		rho=1.0	
		rho=1.2	
GDP=8 minutes	Carrier Delay=0 minute	rho=0.8	
		rho=0.9	
		rho=1.0	
		rho=1.2	
	Carrier Delay=5 minutes	rho=0.8	
		rho=0.9	
		rho=1.0	
		rho=1.2	
	Carrier Delay=10 minutes	rho=0.8	
		rho=0.9	
		rho=1.0	
		rho=1.2	
GDP=15 minutes	Carrier Delay=0 minute	rho=0.8	
		rho=0.9	
		rho=1.0	
		rho=1.2	
	Carrier Delay=5 minutes	rho=0.8	
		rho=0.9	
		rho=1.0	
		rho=1.2	
	Carrier Delay=10 minutes	rho=0.8	
		rho=0.9	
		rho=1.0	
		rho=1.2	

8.2 Outputs of the Case Study

8.2.1 Airport Generated Delay from the Worst Case Scenario

The worst case scenario from the experiment design described in previous section is the condition where there is a 15-mintue GDP, 10-minute Carrier Delay and Departure Demand Ratio is as high as 1.2. The estimated Generated Delay for each airport is listed in Table **8.3**. These delays at the peak operations period at 6PM are assumed due to GDP, Carrier Delay, and high value of Departure Demand Ratio. The airports are listed in the order of their average Airport Delay during summer 2005.

Table **8.3**: Airport Generated Delay Estimated from the Worst Scenario in Case Study Designed for 6 PM. 15-minute GDP, 10-minute Carrier Delay and the Departure Demand Ratio is 1.2.

Airport	PHL	JFK	EWR	ORD	LGA	ATL	DEN
Delay (minute)	45.9	55.4	46.9	38.7	31.5	31.3	24.6

8.2.2 Case Study Result

The combination of settings for GDP Holding Time, Carrier Delay and Departure Demand Ratio results in 48 scenarios (4x3x4). Figure **8.1** shows plots of the Airport Generated Delay. The y-axes are the Predicted Generated Delay in minutes. The x-axes are the different settings for Departure Demand Ratio. The curves with different colors and styles in an airport plot are the Generated Delays given different settings of GDP Holding Time. The plots in the first row are for the scenarios that the Carrier Delay is 10 minutes. The plots in the second row are for the scenarios that the Carrier Delay are 5 minutes, and the plots in the bottom row associated with 0 Carrier Delay.

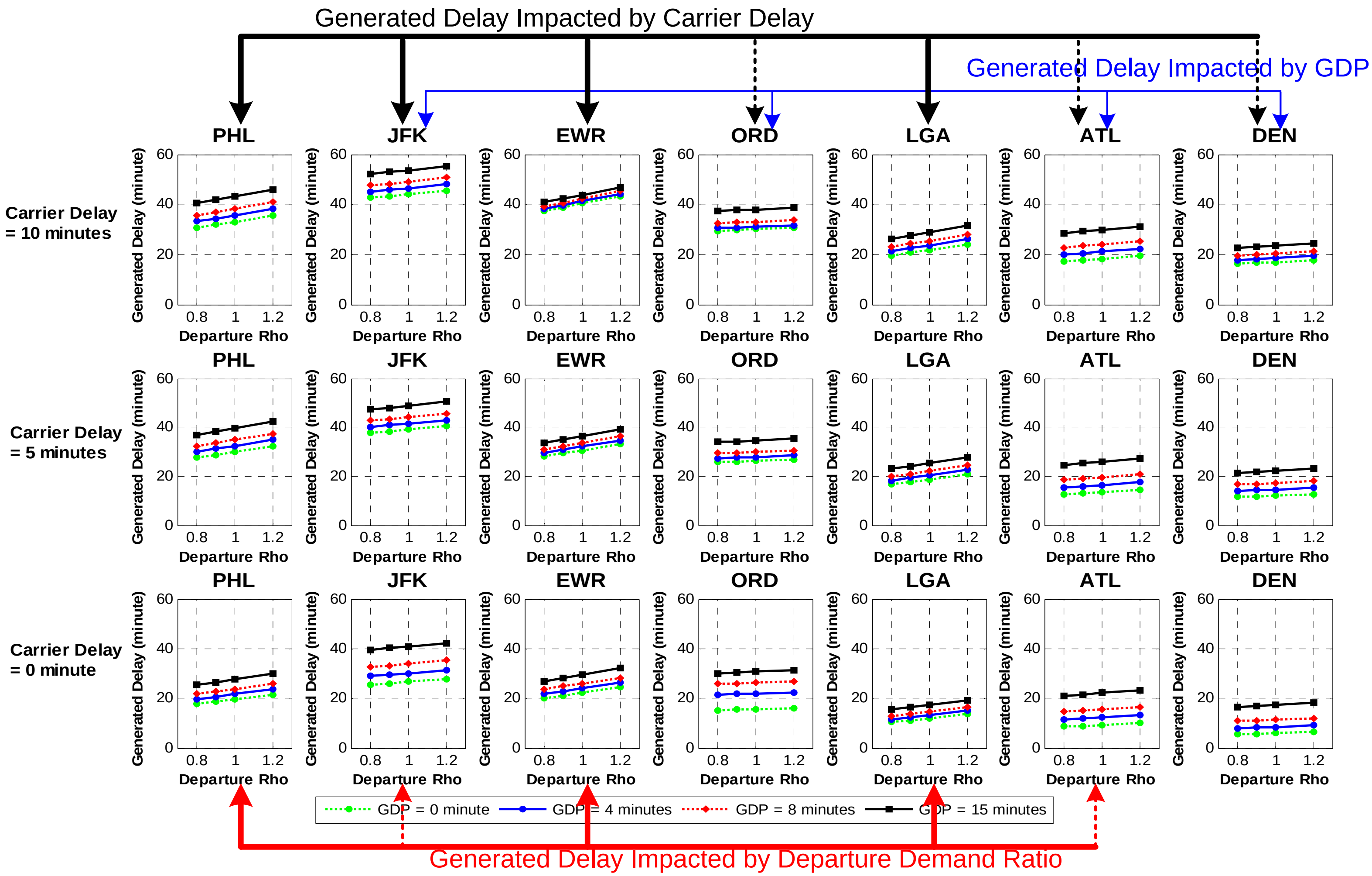


Figure **8.1**: Generated Delays at Each Airport from Case Study Scenarios

Table **8.4** gives a summary of the predicted Generated Delays when the values of inputs of interest were reduced one by one. These values can be found in Figure **8.1**.

8.2.2.1 Reduced Delay Due to the Changes of Departure Demand Ratio

As can be seen from Figure **8.1**, the curves in each plot are almost parallel. Therefore, the impact of Departure Demand Ratio can be approximated by the difference of Generated Delay when reducing Departure Demand Ratio from 1.2 to 0.8 given Carrier Delay is 10 minutes and GDP Holding Time is 15 minutes.

Table **8.4**: Predicted Generated Delay (minute) from each Scenario

	Predicted Generated Delay from the Inputs:				
	GDP=15min CarrierDelay=10min Departure ρ=1.2 DepartureTime=6pm	reduce ρ:1.2 to 0.8	reduce ρ:1.2 to 0.8 Carrier:10 to 0	reduce ρ:1.2 to 0.8 Carrier: 10 to 0 GDP: 15 to 0	reduce ρ:1.2 to 0.8 Carrier: 10 to 0 GDP:15 to 0 Time:6pm to 6am
PHL	46	**40(-5)**	**25(-15)**	18(-8)	**4(-14)**
JFK	55	52(-3)	**40(-13)**	**26(-14)**	**9(-17)**
EWR	47	**41(-6)**	**27(-14)**	20(-7)	**6(-14)**
ORD	39	37(-2)	30(-7)	**15(-15)**	7(-8)
LGA	31	**26(-5)**	**15(-11)**	10(-5)	5(-5)
ATL	31	28(-3)	21(-8)	**9(-12)**	0(-8)
DEN	25	23(-2)	17(-6)	**6(-11)**	6

The third column to the left in Table **8.4** gives the predicted Generated Delay and the reduced delay (in parentheses). Reducing Departure Demand Ratio by 0.4 can reduce Generated Delay most at PHL, EWR, and LGA with 5 minutes or more, and it can only reduce Generated Delay at DEN by 2 minutes. ATL and JFK are in the middle with 3 minutes.

8.2.2.2 Reduced Delay Due to the Changes of Carrier Delay and/or GDP Holding Time

The forth column to the left in the Table **8.4** also gives the reduction of Generated Delay by reducing Carrier Delay given that GDP Holding Time is 15 minutes. Reducing Carrier Delay by 10 minutes can reduce Generated Delay by more than 10 minutes at PHL, JFK, EWR, and LGA.

Reducing GDP Holding Time by 15 minutes given Carrier Delay is 0 can reduce the predicted Generated Delay by more than 10 minutes at JFK, ORD, ATL, and DEN (see the second column to the right in the Table **8.4**). However, this type of calculation of the impact of the Carrier Delay and the GDP Holding Time ignores the interaction between these two factors.

To investigate the impact of Carrier Delay and GDP Holding Time while considering their interrelationship, the Departure Demand Ratio is fixed at 0.8. The Airport Generated Delay was predicted from the 12 scenarios combining 4 different values of GDP Holding Time and 3 values of Carrier Delay: 0, 5, and 10 minutes.

Reducing Carrier Delay

In Figure **8.2**, the x-axis is the Carrier Delay. The y-axis is the predicted Generated Delay. The 4 different colors and styles represent different settings of GDP Holding Time. The predicted Generated Delays in the 4th column in Table **8.4** are the delays of the most left point on the black curve with squares in Figure **8.2**, and the values in the parentheses are the difference between the lowest delay and the highest delay on each of the black curves in Figure **8.2**. Reducing Carrier Delay by 10 minutes can decrease the Generate Delay by more than 10 minutes at PHL, JFK, EWR, and LGA, and by more than 5 minutes at ORD, ATL, and DEN.

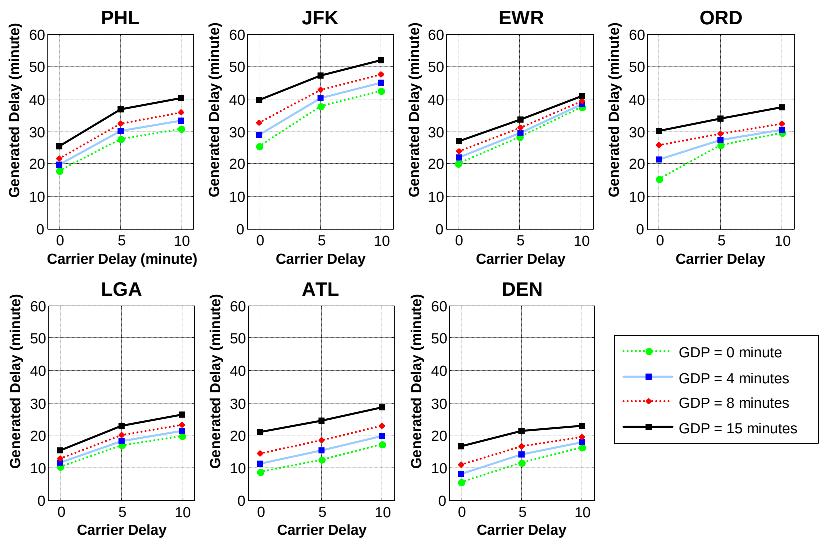


Figure **8.2**: Estimated Airport Generated Delay (minute) vs. Carrier Delay. The Departure Demand Ratio is set at 0.8.

The clear non-linear relationship between Carrier Delay and Airport Generated Delay can be seen from Figure **8.2**, especially at PHL, JFK, ORD, LGA, and DEN. The downward bending of these curves illustrates that the Generated Delay can be reduced more by decreasing the Carrier Delay from 5 minutes to 0 than by decreasing the Carrier Delay from 10 minutes to 5 minutes.

Reducing GDP Holding Time

In Figure **8.3**, the x-axis represents GDP Holding Time. The y-axis represents the predicted Generated Delay. The different colors and styles represent different values of Carrier Delay. This graph evaluated the impact of GDP Holding Time. The delays in the 5th column in Table **8.4** are the lowest delays on the black curve in Figure **8.3** which is associated with 0 Carrier Delay and 0 GDP Holding Time. Reducing GDP Holding time by 15 minutes can reduce Generated Delay by more than 10 minutes at JFK, ORD, ATL, and DEN. Figure **8.3** also shows the non-linear relationship between Generated Delay and GDP Holding Time.

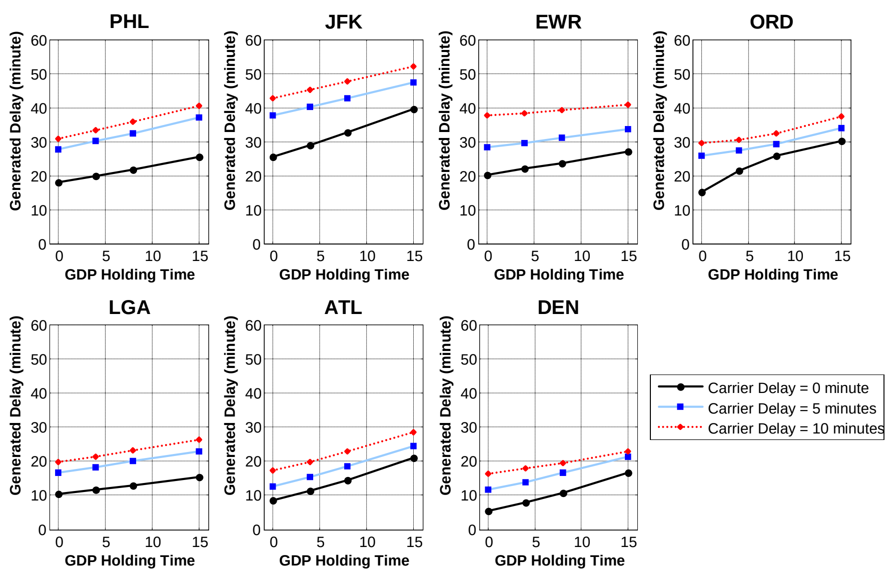


Figure **8.3**: Estimated Airport Generated Delay (minute) vs. GDP Holding Time. The Departure Demand Ratio is set at 0.8.

8.3 Summary of Case Study Results

The selected 7 airports are representatives of different classes of airports. DEN is the largest international airport in the United States by geographic space and number of runways; however, it still has an average of 13 minutes Generated Delay per flight. Analyzing the policy of alleviating delays at these airports is of particular interest.

The conclusions of the case study are summarized below. These conclusions apply only for the range of input data in which the model is valid.

1. Scheduled Departure Time (Time of Day)

At 6 PM, even though there is no GDP, no Carrier Delay and Departure Demand Ratio is as low as 0.8, the Generated Delays at PHL, JFK, EWR, and ORD are 15 minutes or higher. Changing Departure Time from 6 PM to 6 AM, the Generated Delays all fall below 10 minutes at these airports (see last column of Table **8.4**). Therefore, Airport Generated Delay is related to the Time of Day.

2. Departure Demand Ratio

Reducing Departure Demand Ratio from 1.2 to 0.8 can mitigate Generated Delay by more than 5 minutes per flight at PHL, EWR, and LGA. It can reduce Generated Delay by about 3 minutes at ATL and JFK. Making the same change on Departure Demand Ratio at ORD and DEN, which both have 6 runways, has marginal impact on its Generated Delay (2 minutes). The Generated Delay model for ORD has the Runway

Configuration as one of the inputs. In this case study, it was set at the configuration associated with the lowest taxi-out delay. The low impact of Departure Demand Ratio is probably due to its interrelationship with Runway Configuration.

There are two ways to reduce Departure Demand Ratio. One way is to reduce departure demand. The other one is to increase departure capacity.

The departure demand is the result of airline's departure scheduling decisions. It may be controlled by FAA through slot control in the future. The actual departure capacity is not the declared Airport Departure Rate (ADR) by FAA. The capacity-limit factors, such as runways, gates, weather, and some environmental limitations, can diminish the airport capacity, which results in an actual departures rate below the ADR. There are also many cases where the actual departure throughput is above the declared ADR. If over-ADR departure occurs, the ATC controllers have made tradeoffs between departures and arrivals. The actual operation throughput (total departures and arrivals) shall not exceed the summation of AAR and ADR.

Figure **8.4** to Figure **8.8** show the relationship between departure throughput vs. ADR, operation throughput vs. the summation of AAR and ADR, and departure demand vs. ADR using ASPM data for summer 2005 from 6am to midnight. Figure **8.4** to **8.9** show these relationships in scatter plots and probability distribution at EWR, LGA, ORD, ATL and JFK. Figure **8.9** shows JFK in summer 2007. The diagonal lines on the left side represent the situations that ADR equals to departure throughput and AAR+ADR equals to operation throughput. A random jitter was included so data do not overlay. The vertical lines in light color on the right side separate the situations where the throughput is over the declared capacity or the departure demand is over declared capacity.

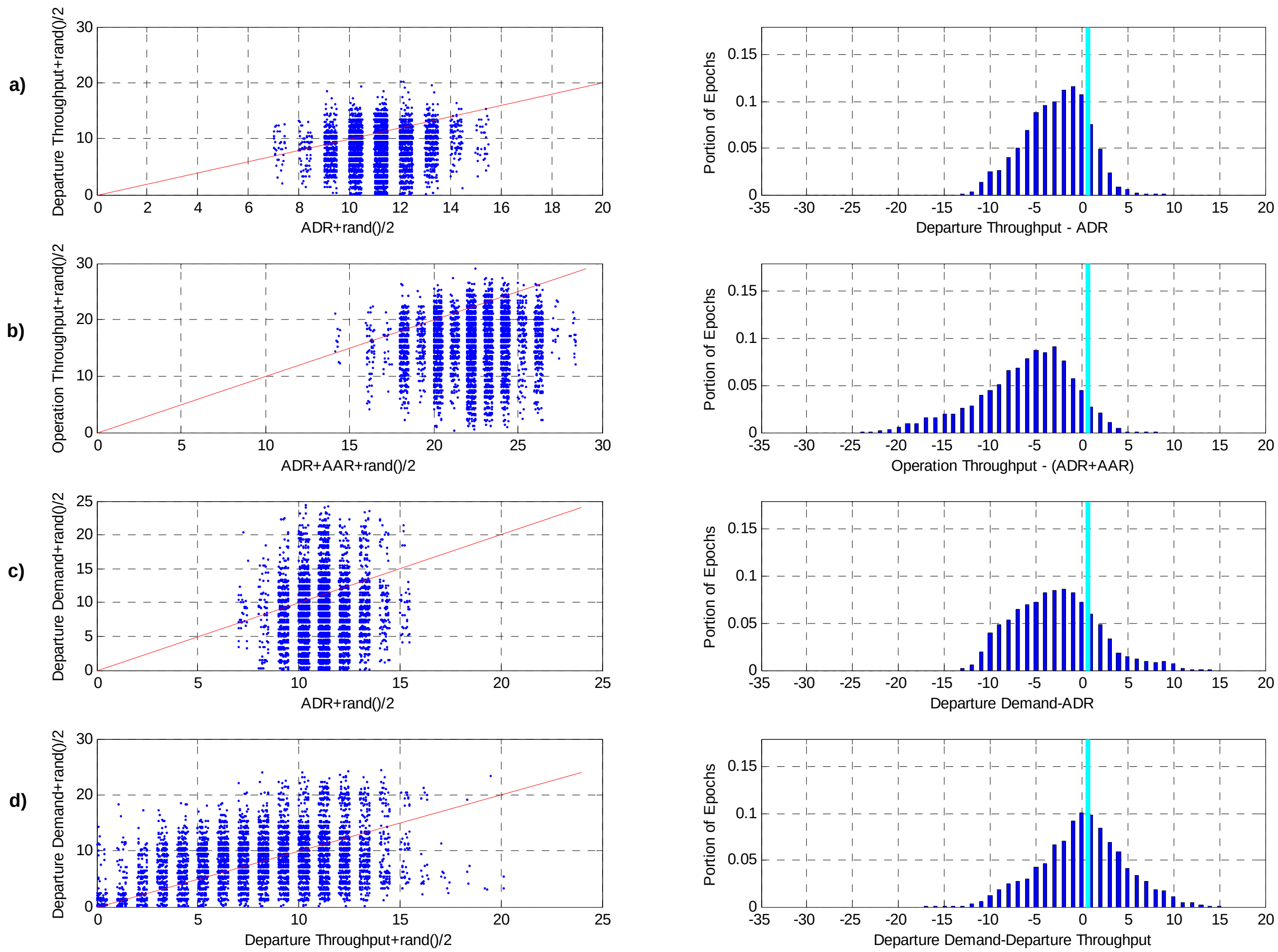


Figure **8.4**: Probability Distribution for EWR, summer 2005 6 am to midnight, a) Departure Throughput vs. ADR, b) Total Operations vs. AAR+ADR, c) Scheduled Departure Demand vs. ADR, d) Scheduled Departure Demand vs. Departure Throughput

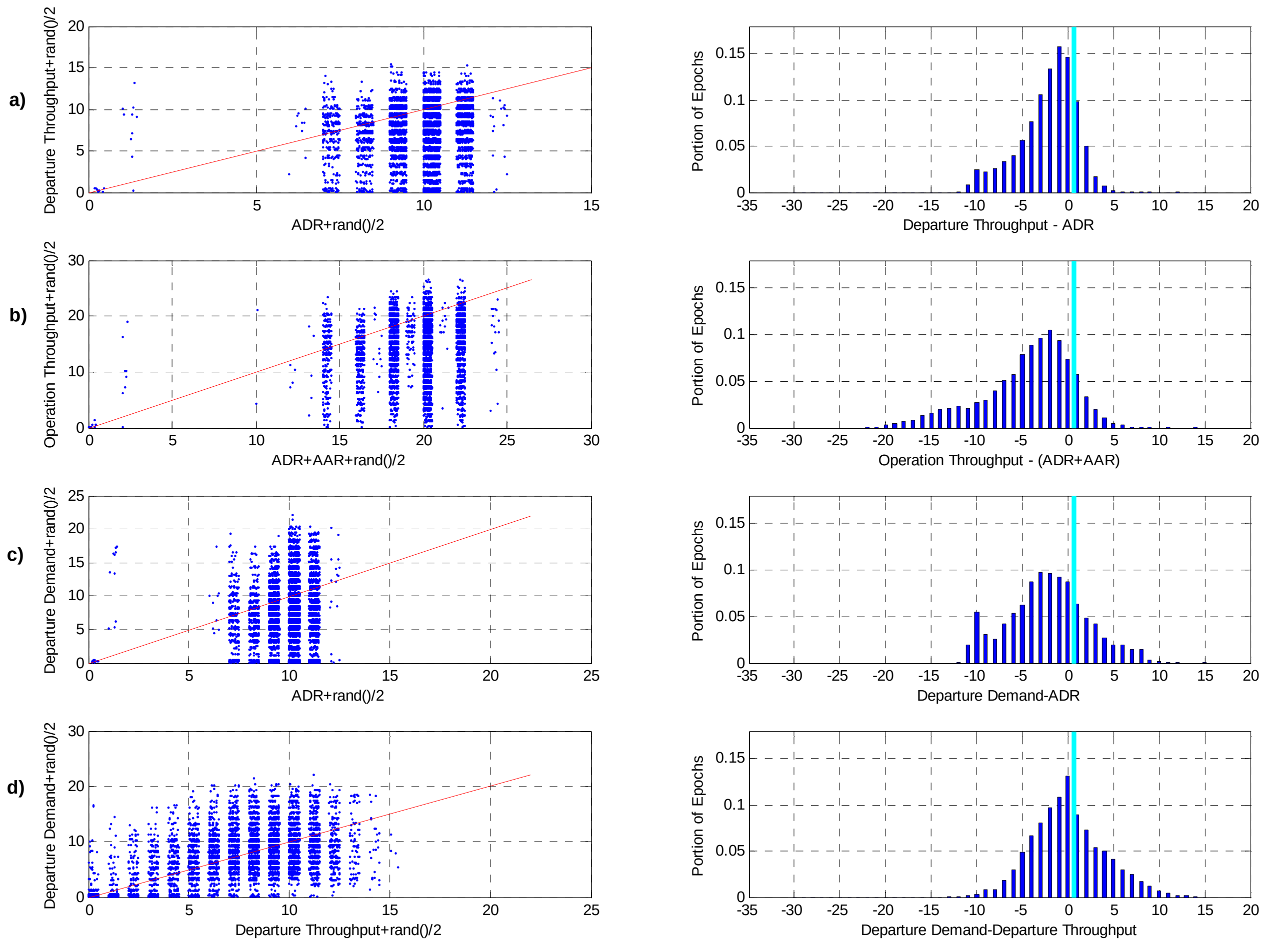


Figure **8.5**: Probability Distribution for LGA, summer 2005 6 am to midnight, a) Departure Throughput vs. ADR, b) Total Operations vs. AAR+ADR, c) Scheduled Departure Demand vs. ADR, d) Scheduled Departure Demand vs. Departure Throughput

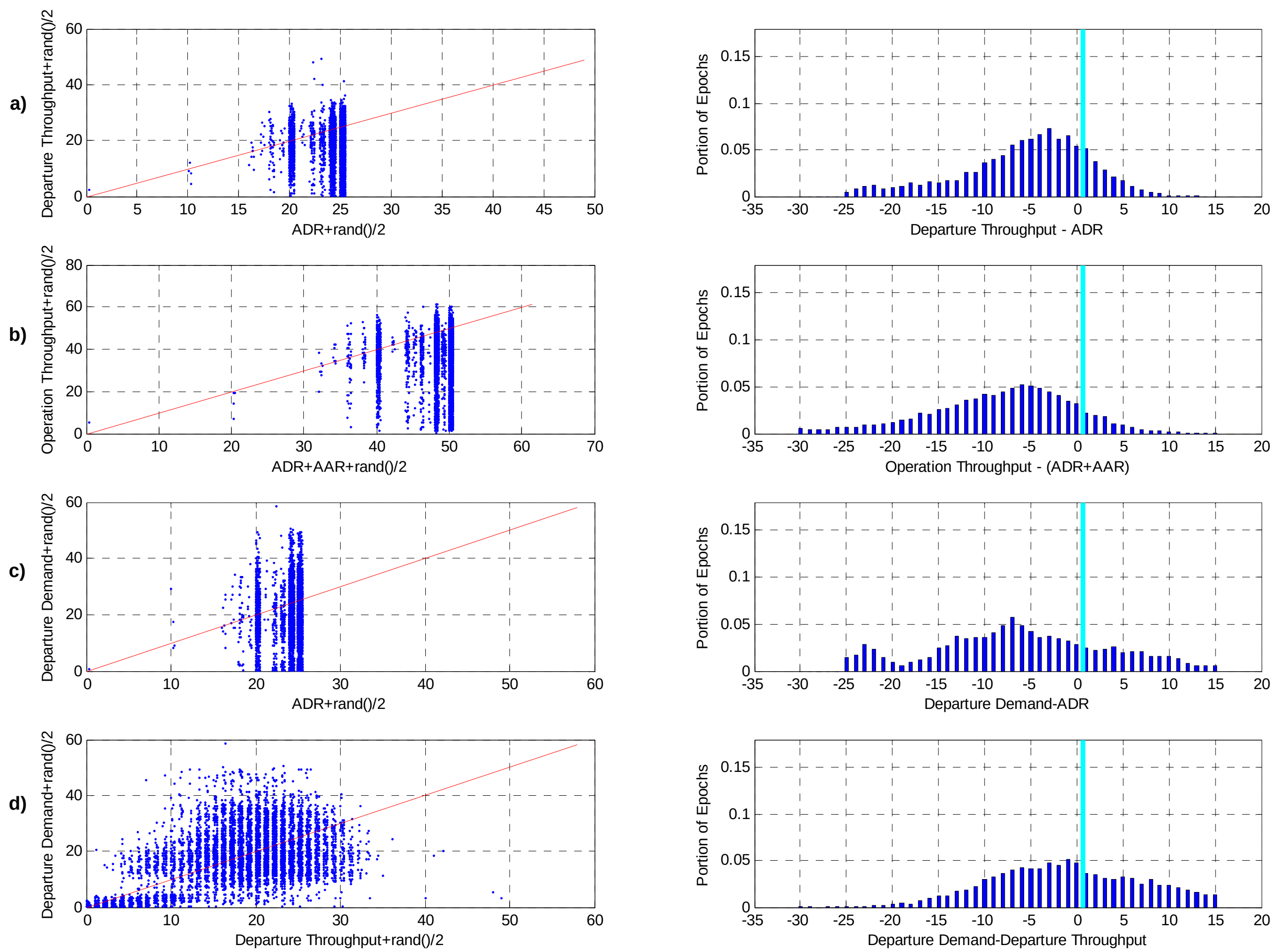


Figure **8.6**: Probability Distribution for ORD, summer 2005 6 am to midnight, a) Departure Throughput vs. ADR, b) Total Operations vs. AAR+ADR, c) Scheduled Departure Demand vs. ADR, d) Scheduled Departure Demand vs. Departure Throughput

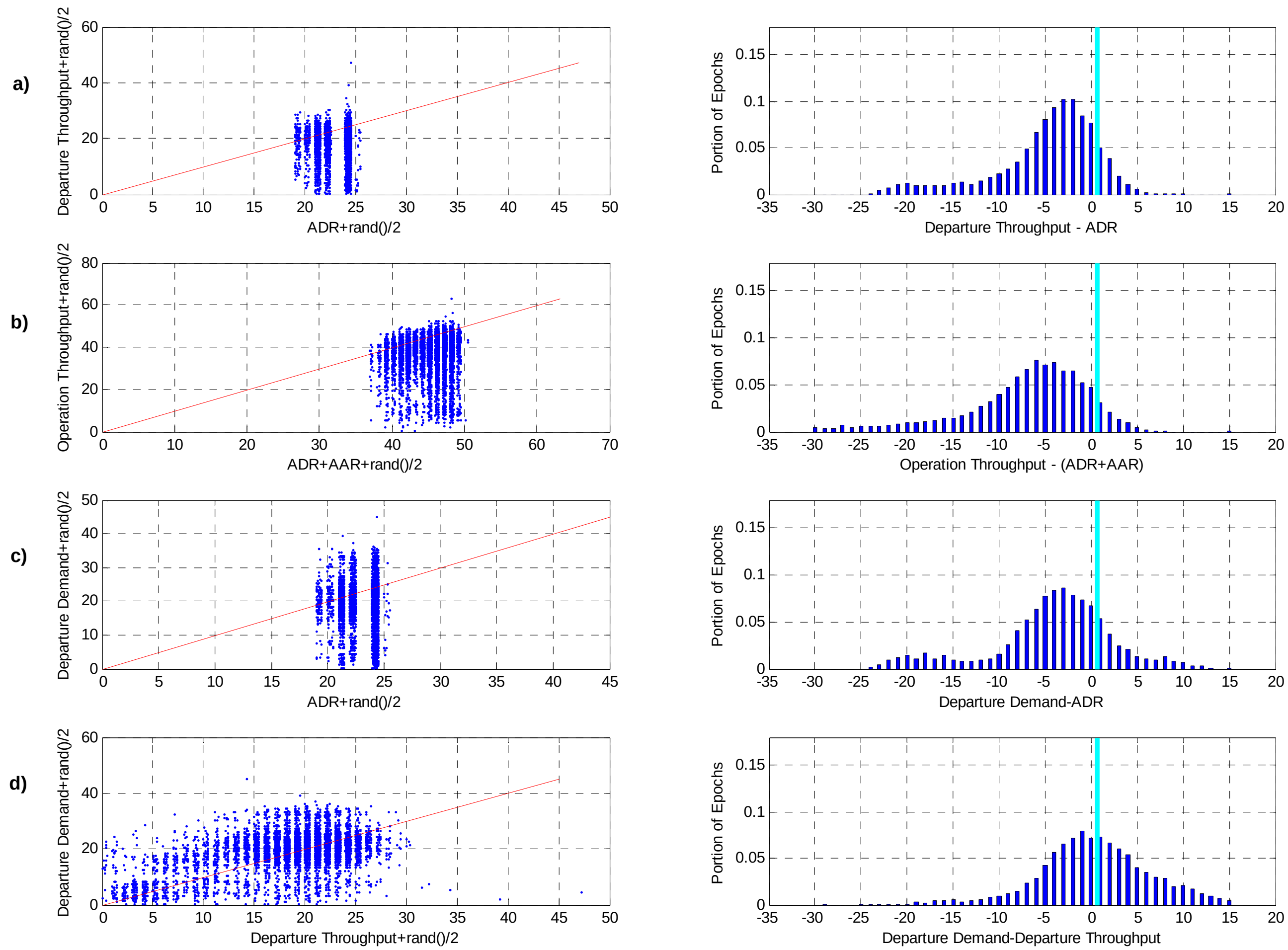


Figure **8.7**: Probability Distribution for ATL, summer 2005 6 am to midnight, a) Departure Throughput vs. ADR, b) Total Operations vs. AAR+ADR, c) Scheduled Departure Demand vs. ADR, d) Scheduled Departure Demand vs. Departure Throughput

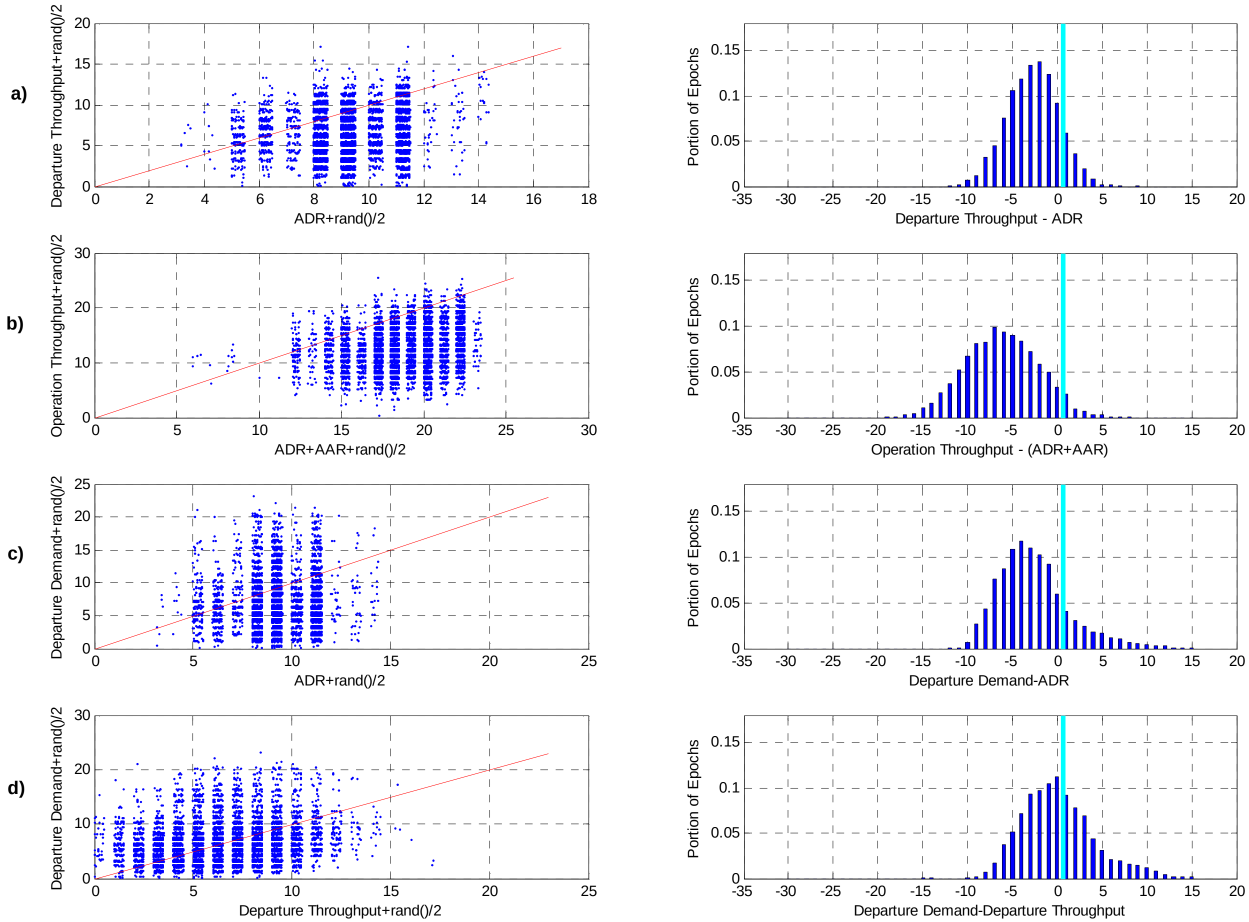


Figure **8.8**: Probability Distribution for JFK, summer 2005 6 am to midnight, a) Departure Throughput vs. ADR, b) Total Operations vs. AAR+ADR, c) Scheduled Departure Demand vs. ADR, d) Scheduled Departure Demand vs. Departure Throughput

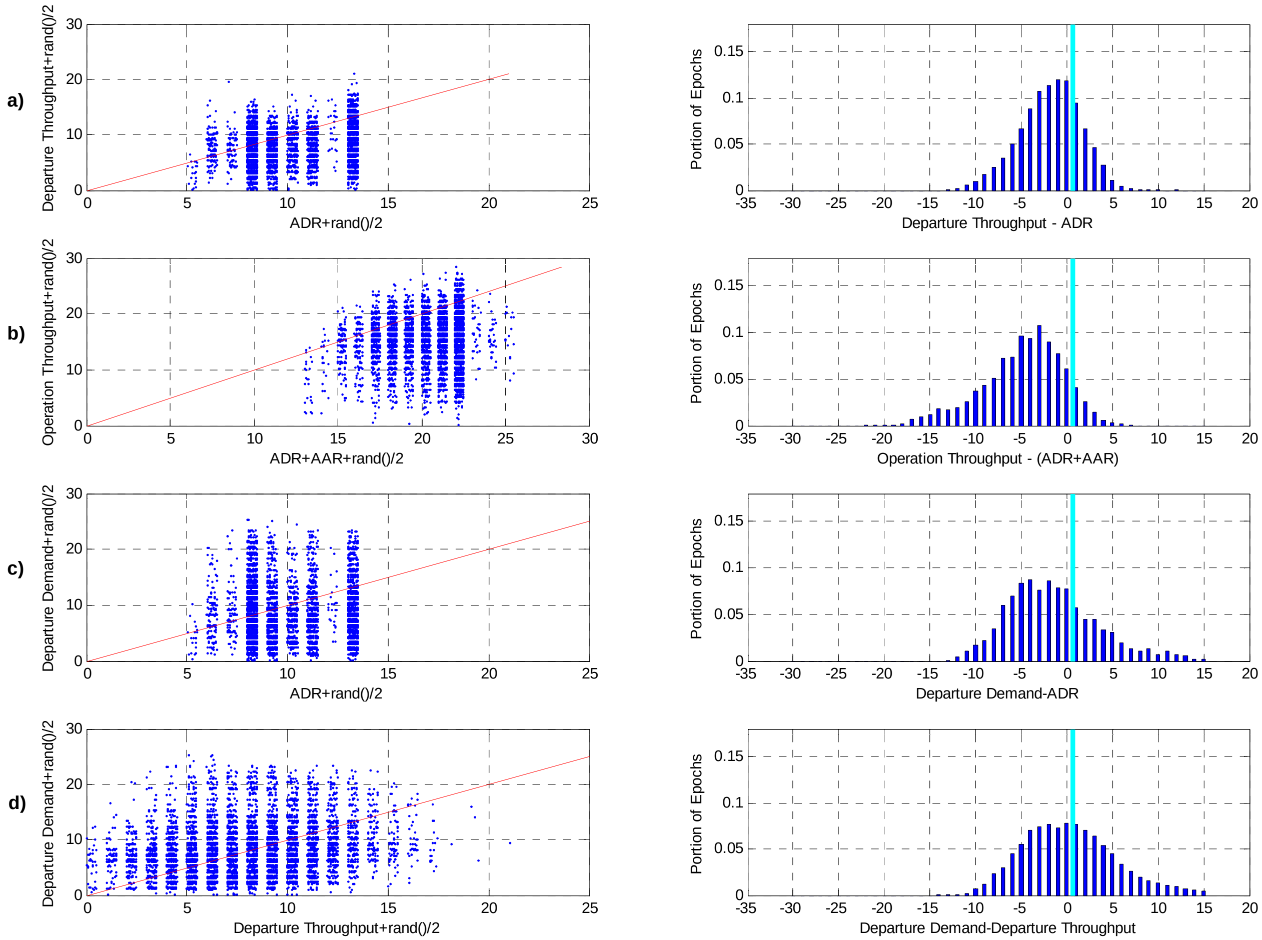


Figure **8.9**: Probability Distribution for JFK, summer 2007 6 am to midnight, a) Departure Throughput vs. ADR, b) Total Operations vs. AAR+ADR, c) Scheduled Departure Demand vs. ADR, d) Scheduled Departure Demand vs. Departure Throughput

Table **8.5** summarizes the accumulated distributions on the right side of Figure **8.4** to Figure **8.9**.

Table **8.5**: Percentages of Actual Operation (departure or departure+arrival) over Declared Capacity (ADR or ARD+AAR), and percentages of the relationship between departure demand and ADR, between departure demand and departure throughput in summer 2005 from 6am to midnight.

Summer 2005 from ASPM	**EWR**	**LGA**	**ORD**	**ATL**	**JFK**	**JFK 2007**
Departure Throughput > ADR	16.4%	17.5%	18.1%	12.7%	12.4%	25.1%
Departure Throughput < ADR	73.1%	67.8%	76.5%	79.7%	78.4%	63.2%
Throughput (Departure + Arrival) > AAR+ADR	6.5%	12.8%	10.3%	8.1%	4.9%	9.1%
Throughput (Departure + Arrival) < AAR+ADR	89.1%	79.8%	86.5%	87.3%	91.8%	84.8%
Departure Demand > ADR	22.3%	25.5%	26.1%	20.1%	17.5%	29.8%
Departure Demand = ADR	7.1%	8.6%	2.8%	6.7%	5.9%	7.7%
Departure Demand < ADR	70.6%	65.8%	71.1%	73.2%	76.6%	62.5%
Departure Demand > Departure Throughput	46.6%	40.5%	43.4%	48.7%	41.1%	45.6%
Departure Demand = Departure Throughput	10.0%	13.0%	4.7%	7.1%	11.1%	7.7%
Departure Demand < Departure Throughput	43.4%	46.5%	51.9%	44.1%	47.8%	46.7%

The airports, EWR and LGA, where the Generated Delays are more sensitive to the Departure Demand Ratio than other airports, were operated over declared departure capacity at more than 15% of time. Departure demand exceeds ADR at these 2 airports more than 20% of time. There is also a large percent of time that the airports are operated under the declared capacity, and departure demand is below the capacity.

Given the current configuration at these airports, it may not be possible to physically increase airport operation capacity, for example adding a new runway. However, there are possibilities that the over-capacity demand can be moved to the time periods when the demand is less than the capacity. Therefore, the ways to reduce Departure Demand Ratio can be either improve airport operation efficiency or to reduce the Departure Demand.

3. GDP Holding Time

From the case study, GDP Holding Time has larger impact on JFK, ORD, ATL, and DEN. Reducing GDP Holding Time by 15 minutes can result in more than 10 minutes reduction of Generated Delay (see Table **8.4**). It has relatively small impact on Generated Delay at PHL, EWR, and LGA when there is carrier delay.

Further investigation of GDP's reveals that ATL, ORD, EWR, LGA and PHL are the top 5 airports issued the highest amount of GDP for their inbound flights in the summer of 2005 (see Table **8.6**). However, the outbound flights from ATL and ORD also have a high chance of being placed under a GDP at their destination. This explains the high influence of GDP Holding Time on Generated Delays at ORD and ATL.

The flights outbound from JFK have very low probability to get assigned in a GDP; however, the Generated Delay of flights leaving JFK is very sensitive to the GDP.

Table **8.6**: Allocation of GDP affected Flights within 34 OEP Airports in Summer 2005

Origin of outbound flights	Percentage of outbound flights affected by GDP	Destination of inbound flights	Percentage of inbound flights affected by GDP
ORD	6.15%	ATL	29.96%
DFW	5.44%	ORD	13.56%
ATL	5.16%	EWR	12.76%
DCA	4.18%	LGA	9.39%
BOS	4.12%	PHL	8.65%
MCO	3.93%	BOS	7.57%
CLT	3.58%	SFO	4.30%
MSP	3.52%	IAD	2.97%
IAH	3.45%	JFK	2.32%
LAX	3.45%	BWI	1.52%
DEN	3.44%	LAS	1.15%
DTW	3.39%	DCA	0.91%
LGA	3.25%	IAH	0.73%
FLL	3.11%	MDW	0.72%
EWR	3.02%	DTW	0.62%
MIA	2.98%	MSP	0.45%
PHL	2.92%	CVG	0.34%
IAD	2.85%	CLT	0.32%
PHX	2.69%	FLL	0.30%
CVG	2.61%	PHX	0.28%
TPA	2.48%	SEA	0.22%
PIT	2.46%	DEN	0.20%
CLE	2.43%	CLE	0.19%
LAS	2.43%	LAX	0.16%
SFO	2.32%	DFW	0.10%
MDW	2.18%	MIA	0.08%
BWI	2.07%	MCO	0.07%
SEA	1.81%	STL	0.05%
MEM	1.66%	PIT	0.04%
STL	1.66%	SAN	0.04%
SLC	1.63%	MEM	0.01%
JFK	1.43%	TPA	0.00%
SAN	1.42%		
PDX	0.75%		
Total	100%	Total	100%

4. Carrier Delay

Reducing Carrier Delay can mitigate Generated Delay at all 7 airports. It has larger impact on the top 3 airports with the worst Generated Delay, PHL, JFK, and EWR. It has big impact on Generated Delay at LGA too. It can also decrease average Generated Delay per flight at ORD, ATL and DEN by more than 5 minutes if it is reduced by 10 minutes.

Overall, these conclusions are only valid for the data within the range of this case study since the models are non-linear and the factor's impacts are expected to be different within different range.

Considering the uncertainty associated with the predictions for the scenarios, the 68% confidence interval for each point estimate is very wide (in Table **8.7**). In Table **8.7**, G represents the square root of Generated Delay and $\hat{\sigma}_{G^2} = 2\hat{\sigma}_G \hat{g}$. Given so much uncertainty in the prediction, the point estimate is very rough for the real delay.

Table **8.7**: Predicted Generated Delay (minute) and 68% Confidence Interval for Each Scenario

		Predicted Airport Generated Delay ($\hat{g}^2$), Standard Deviation ($\hat{\sigma}_{G^2}$), 68% Confidence Interval of Prediction ($\hat{g}^2 \pm \hat{\sigma}_{G^2}$), and Redu. (reduced Generated Delay due to reducing the value of factors)																		
		Departure ρ=1.2 CarrierDelay=10min GDP=15min DepartureTime:6pm			Reduce ρ:1.2 to 0.8				Reduce ρ:1.2 to 0.8 Carrier:10 to 0				Reduce ρ:1.2 to 0.8 Carrier: 10 to 0 GDP: 15 to 0				Reduce ρ:1.2 to 0.8 Carrier: 10 to 0 GDP:15 to 0 Time:6pm to 6am			
	$\hat{\sigma}_G$	$\hat{g}^2$	$\hat{\sigma}_{G^2}$	$\hat{g}^2 \pm \hat{\sigma}_{G^2}$	$\hat{g}^2$	$\hat{\sigma}_{G^2}$	$\hat{g}^2 \pm \hat{\sigma}_{G^2}$	Redu.	$\hat{g}^2$	$\hat{\sigma}_{G^2}$	$\hat{g}^2 \pm \hat{\sigma}_{G^2}$	Redu.	$\hat{g}^2$	$\hat{\sigma}_{G^2}$	$\hat{g}^2 \pm \hat{\sigma}_{G^2}$	Redu.	$\hat{g}^2$	$\hat{\sigma}_{G^2}$	$\hat{g}^2 \pm \hat{\sigma}_{G^2}$	Redu.
PHL	1.6	**46**	22	(24, 68)	**40**	20	(20, 60)	-5	**25**	16	(9, 41)	-15	**18**	14	(4, 32)	-8	**4**	6	(-2, 10)	-14
JFK	1.5	**55**	22	(33, 77)	**52**	22	(30, 74)	-3	**40**	19	(21, 59)	-13	**26**	15	(11, 41)	-14	**9**	9	(0, 18)	-17
EWR	1.7	**47**	23	(24, 70)	**41**	22	(19, 63)	-6	**27**	18	(9, 45)	-14	**20**	15	(5, 35)	-7	**6**	8	(-2, 14)	-14
ORD	1.1	**39**	14	(25, 53)	**37**	13	(24, 50)	-2	**30**	12	(18, 42)	-7	**15**	9	(6, 24)	-15	**7**	6	(1, 13)	-8
LGA	1.6	**31**	18	(13, 49)	**26**	16	(10, 42)	-5	**15**	12	(3, 27)	-11	**10**	10	(0, 20)	-5	**5**	7	(-2, 12)	-5
ATL	1	**31**	11	(20, 42)	**28**	11	(17, 39)	-3	**21**	9	(12, 30)	-8	**9**	6	(3, 15)	-12	**0**	0		-8
DEN	1.1	**25**	11	(14, 36)	**23**	11	(12, 34)	-2	**17**	9	(8, 26)	-6	**6**	5	(1, 11)	-11	**6**	5	(1, 11)	

CHAPTER 9

CONCLUSIONS AND FUTURE WORK

Most major airports in the U.S. are operating near capacity, and the rapid growth of air traffic is causing more delay problems. Based on the delay statistics in Table **A.1**, for a 100-minute inbound delay at the destination, more than 90 minutes of delay occurs at the airports. Even though a large amount of delays (59.1 minutes) were absorbed during the airborne phase, the total Airport Delay propagated in the system through multiple legs of an aircrafts' itinerary still results in a similar amount of total Arrival Delay at outbound destination (Airport Delay 98.6 minutes vs. Total Arrival Delay at Outbound Destination 100 minutes in Figure **1.6**). This motivates the research conducted in this dissertation to model the factors that determine the Airport Delays.

This research developed a method for deriving models to predict airport delays. Through well defined steps of variable selection, six factors out of more then 50 potential factors are shown to be statistically significant for the Airport Generated Delay at most 34 OEP airports.

- Departure Demand Ratio at 30-minute time Window,
- Carrier Delay,
- GDP Holding Time,

- Airline Swap Aircraft Rate,
- Inbound Delay,
- Departure Time.

For Airport Absorbed Delay, 3 factors are indentified to be significant predictors.

- Scheduled Turn-around Time,
- Inbound Delay and
- Carrier Delay

This research confirms quantitatively that the factors which have been considered as causes for delays by previous research and expert operators in the field are contributing factors to airport delay. These factors account for an average of 63% of the variance of Airport Generated Delay and an average of 50% of the variance of Airport Absorbed Delay (average R^2 from Table **6.2**).

The validation results of data samples from 2005 and 2006 prove that airport models developed in this dissertation can provide useful predictions in terms of accuracy. These models also were accurate one year later (see Chapter 6). Estimated on the hold out sample, the average absolute prediction error is 6.2 minutes of 34 OEP airports for 15 days in Aug. 2005 with minimum 4.2 minutes for LAX and maximum 9.2 minutes for PHL (see Table **6.2**). The mean absolute prediction error for data in August 2006 is below 10 minutes at 31 airports. The maximum error is 12.4 minutes for PHL (see Table **6.3**).

With such accuracy, the models developed in this research can not only be used to predict airport delay, to analyze the impact of delay mitigation policy on an airport, but also are ideal building blocks for NAS simulation models to identify the system impact of delay mitigation policies.

9.1 Conclusions from Sensitivity Analysis and Case Study

An approximate sensitivity analysis was performed on the validated models to provide quantitative measures of the influence of important factors. The summaries of analysis results are as follows.

Airports are unique. Different airport should implement different delay mitigation strategies.

1. **The factors influencing Airport Generated Delay and Absorbed Delay are not exactly the same across 34 airports (see Table A.5 and A.6**).

Each airport has its own operation characteristics. The models in Appendix B have different predictors, basis functions and coefficients. Each basis function represents the contribution of a predictor over a certain range of values. The coefficient of each basis function indicates the degree of influence of the corresponding factor when its value is in that range.

Using Departure Time as an example, the airports behave differently at different times of the day. Figure **9.1** shows the break-points (knots) for Departure Time at the airports in New York which have Departure Time in their Generated Delay model.

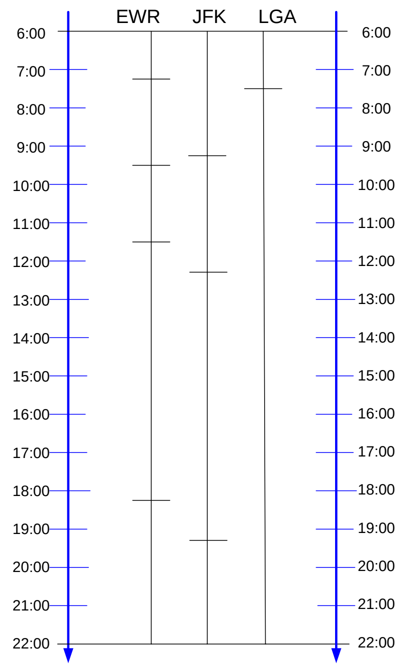


Figure **9.1**: The Break-points of Departure Time in the Airport Generated Delay Models

As can be seen from the table, no two airports have the same break-points of Departure Time. The break points at EWR and JFK are different even though they are geographically close. However, the difference of EWR and JFK is not large. The first group of close break-points is around 9:00 to 10:00 in the morning, which is the time the airports start to get crowded. The second group of break-points is around 12:00 at noon. And, the third group of break-points is around 6pm, which is the time when a bank of flights departs for Europe.

Since the airport's behavior is not the same, the assumption of treating airports as identical (e.g. Boswell and Evans 1997) oversimplifies the problem and should be avoided.

2. The factors' impact on delays are different

The significant factors for airport delays were narrowed down from more than 50 to 6 factors for Airport Generated Delay and 4 factors for Absorbed Delay by following the approaches described in Chapter 4. Among these factors, some of them have more influence than others on the airport delays.

The sensitivity analysis conducted in this research investigated which factors have the most impact on the overall performance. The second column of Table **9.1** shows the means of positive values of significant factors for Airport Generated Delay and Absorbed Delay across 34 airports using the data of summer 2005. These are the same values used for sensitivity analysis in Chapter 7.

Table **9.1**: Summary of Mean Value of Factors at 34 OEP Airports in June and July 2005 and the Average Slopes of these Factors in Sensitivity Analysis. The Slope measures the amount of variation of delays (in minute) given 10% increment of the mean of corresponding factor.

		Average Slopes	
	Mean	Generated Delay	Absorbed Delay
Carrier Delay (minute)	10.35	**2.78**	0.38
GDP Holding Time (minute)	19.22	**2.79**	0.16
Departure Demand Ratio	1.25	**3.86**	
Inbound Delay (minute)	20.65	-0.92	**-2.95**
Swap Aircraft Rate	0.23	0.49	
Schedule Turnaround Time (minute)	60.04		**-3.53**
Number of Seats	134		1.56

The last two columns of Table **9.1** show the average slopes across 34 airports based on the output of sensitivity analysis in Chapter 7. For example, the first entry of the third column is the average value of the slopes reported in Table **7.1**.

The **Carrier Delay, GDP Holding Time,** and **Departure Demand Ratio** are the most significant factors to Airport Generated Delay. The **Scheduled Departure Time** is a very important factor in the model. It is not included in the sensitivity analysis because Scheduled Departure Time is decided by airlines and cannot be manipulated.

Reducing Carrier Delay and GDP Holding Time will reduce Generated Delay at all 34 airports. Departure Demand Ratio appears at 23 Airport Generated Delay models. Reducing Departure Demand Ratio will also reduce Generated Delay.

Inbound Delay and Swap Aircraft Rate only slightly affect Airport Generated Delay. The reason for the small impact of Swap Aircraft Rate is that swapping aircraft only occurs 4% of time (see section 7.2.4). With such small percentage, reducing the possibility of swapping aircraft does not significantly affect the overall performance in 2-month training data period.

The reason for the small impact of Inbound Delay may be due to its colinearity with Scheduled Departure Time. When the major part of the variance of Generated Delay has been explained by Scheduled Departure Time, the Inbound Delay does not seem to affect Generated Delay much (see Figure **7.8**). The negative average slope of impact of

Inbound Delay to Generated Delay implies that implies that airports were forced to generate less delay for flights with longer in-bound delay.

The **Inbound Delay** and **Scheduled Turnaround Time** are the most significant factors to Airport Absorbed Delay. Extended schedule turn-around time results in longer absorbed delay and all airports can absorb more delay when carrier delay and GDP holding time are reduced.

3. For airports that share the same factors, a given factor has different influence at different airports.

The air traffic manager and stakeholder (e.g. ATC) should apply different policy on the airports according to their dominant factors.

Carrier Delay is solely the responsible of airline and changing it will not adversely affect the airlines' market and profit. The causes of these delays are the circumstances considered within the airline's control by FAA. The Airport Generated Delay can be mitigated by reducing Carrier Delay more significantly at some airports than other. These airports include ORD, ATL, MSP, PHL, PHX, DEN, DTW, JFK, EWR, and DFW (see Figure **7.2** and Table **7.1**).

Departure Demand Ratio measures the balance of departure demand and actual departure capacity for each airport.

Both the mean (1.25) and the median (1.05) of Departure Demand Ratio across 34 airports show that the major airports are operating up to their capacity. The near-capacity operation is one of the most significant reasons for delays.

From the outputs of the Sensitivity Analysis in Chapter 7, the Airport Generated Delay at PHL, LGA, EWR, and ATL can be alleviated by the greatest amount by reducing Departure Demand Ratio (see Figure **7.6** and Table **7.4**). This result is similar to the results of Case Study in Chapter 8 Table **8.4**.

There were situations that these airports operated above its designed capacity, there also were situations the airports were operated below its designed capacity (see Figure **8.4** to Figure **8.9** and Table **8.5**). There were situations that the demand at these airports is above its designed capacity, there also were situations that demand is below capacity.

The feasible solutions to reduce Departure Demand Ratio can be either to reduce departure demand or to increase airport capacity. If adding new runways is too costly given the financial or environmental consideration, effectively utilizing the current facility is only solution. There are large percentages of time that the demand is below the capacity (see Table **8.5**). It will mitigate the congestion at the airport if some demand at peak time is moved to other times. The airlines are not willing to sacrifice profit by reducing their operation or changing their schedule. Hence, an enforced slot control or other incentive may be a way to solve this problem.

GDP is issued by the Air Traffic Control system Command Center (ATCSCC). The intention of GDP is to move the airborne delay to outbound destination to ground delay at origin. The reason that the destination airports issue GDP to inbound flights at origin is the arrival demand over arrival capacity. Instead of recommending reducing GDP Holding Time at Origin airport, it would make more sense to reduce the arrival demand at destination.

Although ground delay is safer and cheaper than airborne delay, it causes problems at origin airport and airline operation. Reducing GDP Holding Time can reduce Airport Generated Delay at DFW, ATL, ORD, CLT, DCA, CVG, BOS, IAH, and MIA more than other airports (see Table **7.2)**. There airports are in the top 10 airports which have the highest percentage of GDP to outbound flights, except CVG and MIA (see Table Error! Reference source not found.).

4. At a given airport, the rate of change of airport delay due to a specific variable differs at different ranges of the variable.

Nonlinearity is essential. Based on the outputs from section 8.2.2.2, reducing Carrier Delay from 10 minutes to 5 minutes does not have the same amount of reduction of Generated Delay as reducing the Carrier Delay from 5 to 0 at PHL and JFK. Hence, preventing Carrier Delay at JFK and PHL is more important than reducing Carrier Delay.

Intuitively, longer turn-around time at airport terminal enables aircraft to make up delays. The sensitivity analysis on Scheduled Turn-around Time shows this trend when the time was artificially added or reduced within 30%. However, adding more than 40% turn-around time results in less absorbed delay at some airports (see Figure **7.10**).

Overall, the analysis results suggest that each airport should be treated differently according to its unique characteristics in the process of improving NAS performance. A GUI tool to predict Airport Delay at LGA was developed based on the developed models. The same tools can be developed for other airports as well. Such tools would enable AOC and TFM personnel to perform "what if" analysis by making changes in causal factors at various times of the day and observing the predicted effects. The display would include multiple delay predictions to better understand the impact of one or more types of delays.

9.2 Recommendations for Future Work

There are several limitations of this research.

The first limitation is the limitation of regression analysis. An advantage of regression modeling is the ability to interpolate from regression analysis, i.e., drawing inference about $E(Y \mid X = x)$ for x within the range of observed values of X. However, drawing an inference about $E(Y \mid X = x)$ for x outside the range of the observed values of X can be dangerous.

The data for this study did not include manipulation of the factors influencing delay. For this reason, care must be taken in interpreting the model for policy purposes.

The assumptions required to interpret the relationships causally would have to be carefully justified.

Another limitation is that for a dynamic moving system like NAS, a one-time model development process is not sufficient. The models need to be updated in order to be able to represent the real system. It suggests that a running 60-day window to predict next 30 days to account for adaption of system would be a better strategy for using this model.

The limitation of the approach of sensitivity analysis is that manipulating one factor at a time and keeping other factors unchanged assumes that these factors vary independently from each other. In reality, when changing one factor, the value of other factors may change accordingly. The interrelationships among the factors will be investigated in future research.

This research focuses on delays at airports. The air space is as important as airports in air traffic management. The overall effect of the airborne phase is to absorb system delays rather than to generate delays. One possible reason for this phenomenon is that air traffic control initiatives such as GDP have transferred the delay in air space to the ground. Another possibility is that airlines pad their schedule so that they have more room to make up lost time when they are behind schedule. These hypotheses need to be investigated by future study.

Once the airborne phase can be accurately estimated, the models of airports and airborne phases can be linked together to construct a network model of delay propagation for NAS.

9.3 Published Results

This research has been documented in several papers as follows:

1. Xu, N., L. Sherry, and K.B. Laskey (2007). Sensitivity Analysis of Factors Causing Airport Delay (to be submitted).

2. Xu, N., L. Sherry, and K.B. Laskey (2007). Multi-factor Models for Predicting Delays at U.S. Airports, Transportation Research Board, 2007.

3. Xu, N., K.B. Laskey, C.H. Chen, S.C. Williams and L. Sherry (2007). Bayesian Network Analysis of Flight Delays. Proc. Transportation Research Board 86th Annual Meeting Compendium of Papers CD-ROM, 2007.

4. Laskey, K., N. Xu, and C.H. Chen (2006). Propagation of Delays in the National Airspace System. Proc. of the 22nd Conf. of Uncertainty in Artificial Intelligence, Cambridge, MA. 2006. pp. 265-272.

5. Xu, N., B.K. Laskey,G. Donohue, and C.H. Chen (2005). Estimation of Delay Propagation in the National Aviation System Using Bayesian Networks. Proc. of 6th USA/Europe Air Traffic Management R&D Symposium, Baltimore, MD. 2005.

APPENDIX A

TABLES AND FIGURES

Table **A.1**: Total Delays in Summer 2005 (minutes). The airports are ordered by total arrival delay at outbound destinations.

Airport	Number of outbound flight	**Inbound Delay from pre leg**	**Early Arrival Gap**	Airport Generated Delay	Airport Absorbed Delay	**Airport Delay**	Airborne Generated Delay	Airborne Absorbed Delay	**Airborne Delay**	**Arrival Delay At Outbound Dest.**
ATL	43885	**526888**	**71938**	824013	-218286	**605727**	57655	-394384	**-336729**	**867824**
ORD	45834	**328631**	**107658**	1021362	-196916	**824446**	46401	-538984	**-492583**	**768154**
DFW	35580	**217363**	**68710**	603962	-132085	**471877**	63608	-256300	**-192692**	**565258**
EWR	24341	**247336**	**30934**	612241	-109039	**503202**	19517	-393709	**-374192**	**407280**
PHL	21699	**194939**	**44203**	613518	-67987	**545531**	14475	-401113	**-386638**	**398034**
MSP	22916	**116703**	**35594**	495080	-110194	**384886**	37877	-182047	**-144170**	**393014**
DTW	21961	**120609**	**35203**	417545	-103114	**314431**	41735	-148839	**-107104**	**363138**
DEN	30114	**81674**	**71547**	364923	-112214	**252709**	66872	-161558	**-94686**	**311243**
BOS	25841	**163846**	**42084**	419562	-100875	**318687**	35900	-257769	**-221869**	**302748**
JFK	15958	**144880**	**12851**	432341	-72359	**359982**	18843	-236535	**-217692**	**300021**
LGA	23176	**228904**	**33222**	463935	-124513	**339422**	17996	-333665	**-315669**	**285879**
IAH	24576	**61276**	**53262**	472821	-88858	**383963**	30607	-253148	**-222541**	**275960**
IAD	19052	**131164**	**33805**	383558	-82029	**301529**	23123	-215704	**-192581**	**273917**
CLT	19103	**102143**	**31158**	348464	-77401	**271063**	26424	-171927	**-145503**	**258861**
MCO	21022	**132354**	**34389**	259151	-84675	**174476**	49603	-147053	**-97450**	**243770**
DCA	20386	**99725**	**35657**	304107	-77928	**226179**	31772	-162051	**-130279**	**231281**
PHX	26832	**55443**	**60311**	390254	-82953	**307301**	36942	-228973	**-192031**	**231025**
MIA	13615	**79313**	**14615**	270751	-54424	**216327**	26722	-109026	**-82304**	**227950**
LAX	30921	**90424**	**58990**	321612	-113132	**208480**	64216	-196264	**-132048**	**225845**
LAS	22447	**127865**	**46204**	283206	-76787	**206419**	31084	-189659	**-158575**	**221912**
BWI	16011	**99962**	**31201**	200354	-45414	**154940**	26427	-131761	**-105334**	**180768**

Table **A.1** Total Delays in Summer 2005 (minutes). The airports are ordered by total arrival delay at outbound destinations. (continued)

Airport	Number of outbound flight	**Inbound Delay from pre leg**	**Early Arrival Gap**	Airport Generated Delay	Airport Absorbed Delay	**Airport Delay**	Airborne Generated Delay	Airborne Absorbed Delay	**Airborne Delay**	**Arrival Delay At Outbound Dest.**
SEA	15808	**91970**	**23300**	196234	-67046	**129188**	36128	-103234	**-67106**	**177353**
FLL	14244	**96762**	**21930**	211244	-48768	**162476**	27730	-132672	**-104942**	**176226**
CVG	18280	**30411**	**52496**	260366	-74994	**185372**	30213	-138867	**-108654**	**159624**
CLE	15000	**70445**	**26300**	203108	-66912	**136196**	25379	-103919	**-78540**	**154401**
MDW	16457	**60513**	**52721**	230571	-48645	**181926**	20991	-166906	**-145915**	**149246**
SFO	18302	**83730**	**31898**	211642	-94361	**117281**	37178	-126310	**-89132**	**143777**
TPA	13845	**83222**	**18895**	167675	-50878	**116797**	28024	-103669	**-75645**	**143269**
PIT	12956	**56306**	**26838**	159094	-51282	**107812**	25389	-88122	**-62733**	**128224**
STL	13633	**69001**	**28523**	134412	-40733	**93679**	27173	-94892	**-67719**	**123484**
SLC	16577	**36079**	**31691**	185788	-57107	**128681**	23781	-109596	**-85815**	**110636**
SAN	16653	**72245**	**21071**	146359	-57271	**89088**	28148	-101481	**-73333**	**109071**
MEM	8786	**40293**	**17917**	125526	-38032	**87494**	18651	-56066	**-37415**	**108289**
PDX	8591	**36478**	**11091**	68197	-32062	**36135**	17860	-43869	**-26009**	**57695**
Total	714402	**4178897**	**1318207**	11802976	-2859274	**8943702**	1114444	-6480072	**-5365628**	**9075177**

Table **A.2**: Average Delays per Flight in Summer 2005 (minutes). The airports are ordered by average airport delay.

Airport	**Inbound Delay from pre leg**	**Early Arrival Gap**	Airport Generated Delay	Airport Absorbed Delay	**Airport Delay**	Airborne Generated Delay	Airborne Absorbed Delay	**Airborne Delay**	**Arrival Delay At Outbound Dest.**
PHL	9.0	2.0	28.3	-3.1	25.1	0.7	-18.5	-17.8	18.3
JFK	9.1	0.8	27.1	-4.5	22.6	1.2	-14.8	-13.6	18.8
EWR	10.2	1.3	25.2	-4.5	20.7	0.8	-16.2	-15.4	16.7
ORD	7.2	2.3	22.3	-4.3	18.0	1.0	-11.8	-10.7	16.8
MSP	5.1	1.6	21.6	-4.8	16.8	1.7	-7.9	-6.3	17.2
MIA	5.8	1.1	19.9	-4.0	15.9	2.0	-8.0	-6.0	16.7
IAD	6.9	1.8	20.1	-4.3	15.8	1.2	-11.3	-10.1	14.4
IAH	2.5	2.2	19.2	-3.6	15.6	1.2	-10.3	-9.1	11.2
LGA	9.9	1.4	20.0	-5.4	14.6	0.8	-14.4	-13.6	12.3
DTW	5.5	1.6	19.0	-4.7	14.3	1.9	-6.8	-4.9	16.5
CLT	5.3	1.6	18.2	-4.1	14.2	1.4	-9.0	-7.6	13.6
ATL	12.0	1.6	18.8	-5.0	13.8	1.3	-9.0	-7.7	19.8
DFW	6.1	1.9	17.0	-3.7	13.3	1.8	-7.2	-5.4	15.9
BOS	6.3	1.6	16.2	-3.9	12.3	1.4	-10.0	-8.6	11.7
PHX	2.1	2.2	14.5	-3.1	11.5	1.4	-8.5	-7.2	8.6
FLL	6.8	1.5	14.8	-3.4	11.4	1.9	-9.3	-7.4	12.4
DCA	4.9	1.7	14.9	-3.8	11.1	1.6	-7.9	-6.4	11.3
MDW	3.7	3.2	14.0	-3.0	11.1	1.3	-10.1	-8.9	9.1
CVG	1.7	2.9	14.2	-4.1	10.1	1.7	-7.6	-5.9	8.7
MEM	4.6	2.0	14.3	-4.3	10.0	2.1	-6.4	-4.3	12.3
BWI	6.2	1.9	12.5	-2.8	9.7	1.7	-8.2	-6.6	11.3

Table **A.2**: Average Delays per Flight in Summer 2005 (minutes). The airports are ordered by average airport delay. (continue)

Airport	**Inbound Delay from pre leg**	**Early Arrival Gap**	Airport Generated Delay	Airport Absorbed Delay	**Airport Delay**	Airborne Generated Delay	Airborne Absorbed Delay	**Airborne Delay**	**Arrival Delay At Outbound Dest.**
LAS	5.7	2.1	12.6	-3.4	9.2	1.4	-8.4	-7.1	9.9
CLE	4.7	1.8	13.5	-4.5	9.1	1.7	-6.9	-5.2	10.3
TPA	6.0	1.4	12.1	-3.7	8.4	2.0	-7.5	-5.5	10.3
DEN	2.7	2.4	12.1	-3.7	8.4	2.2	-5.4	-3.1	10.3
PIT	4.3	2.1	12.3	-4.0	8.3	2.0	-6.8	-4.8	9.9
MCO	6.3	1.6	12.3	-4.0	8.3	2.4	-7.0	-4.6	11.6
SEA	5.8	1.5	12.4	-4.2	8.2	2.3	-6.5	-4.2	11.2
SLC	2.2	1.9	11.2	-3.4	7.8	1.4	-6.6	-5.2	6.7
STL	5.1	2.1	9.9	-3.0	6.9	2.0	-7.0	-5.0	9.1
LAX	2.9	1.9	10.4	-3.7	6.7	2.1	-6.3	-4.3	7.3
SFO	4.6	1.7	11.6	-5.2	6.4	2.0	-6.9	-4.9	7.9
SAN	4.3	1.3	8.8	-3.4	5.3	1.7	-6.1	-4.4	6.5
PDX	4.2	1.3	7.9	-3.7	4.2	2.1	-5.1	-3.0	6.7

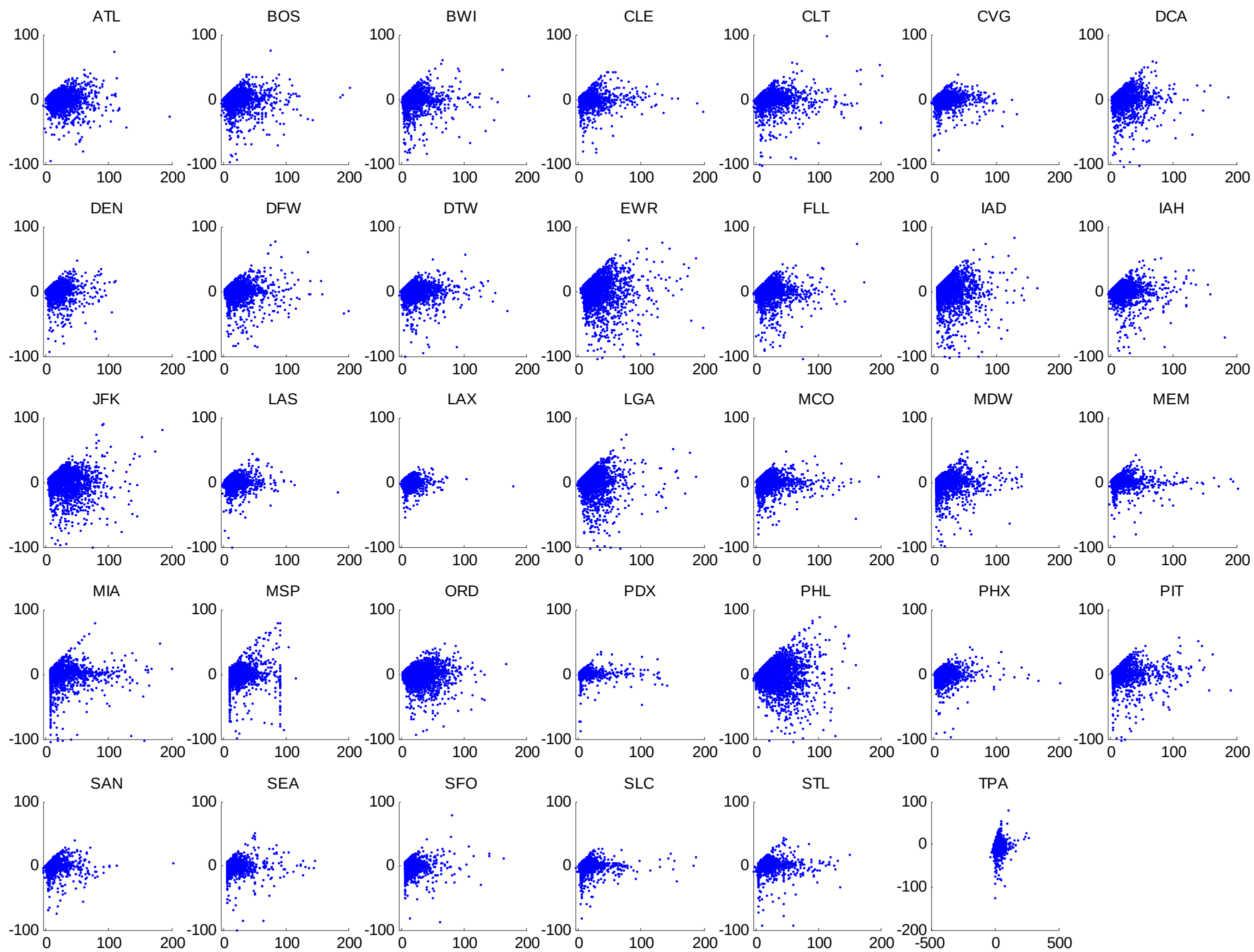


Figure **A.1**: Scatter Plot of Residuals of Generated Delay on Original Scale from Full-size Model.

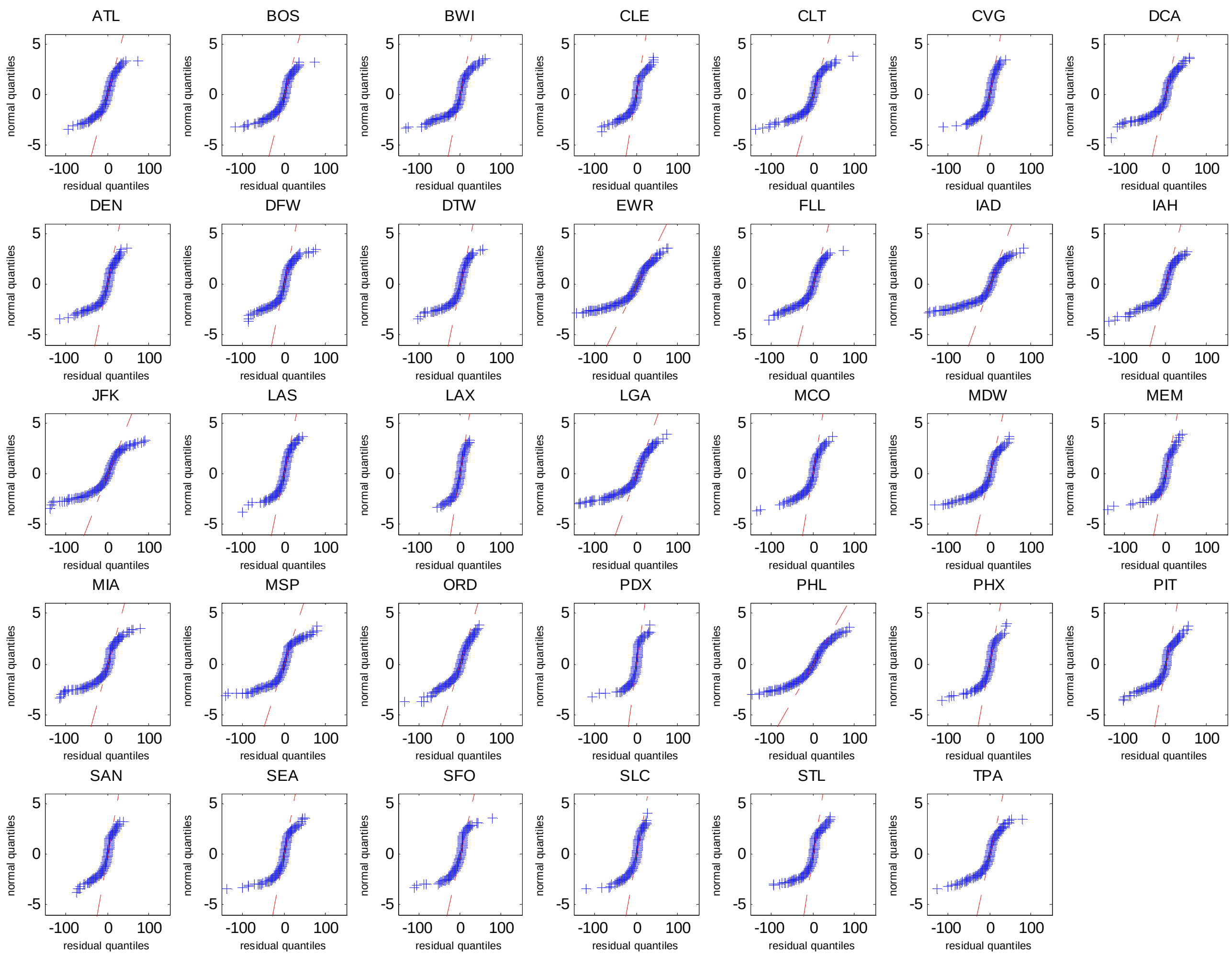


Figure **A.2**: Quantile_Quantile Plot of Residuals of Generated Delay on Original Scale from Full-size Model

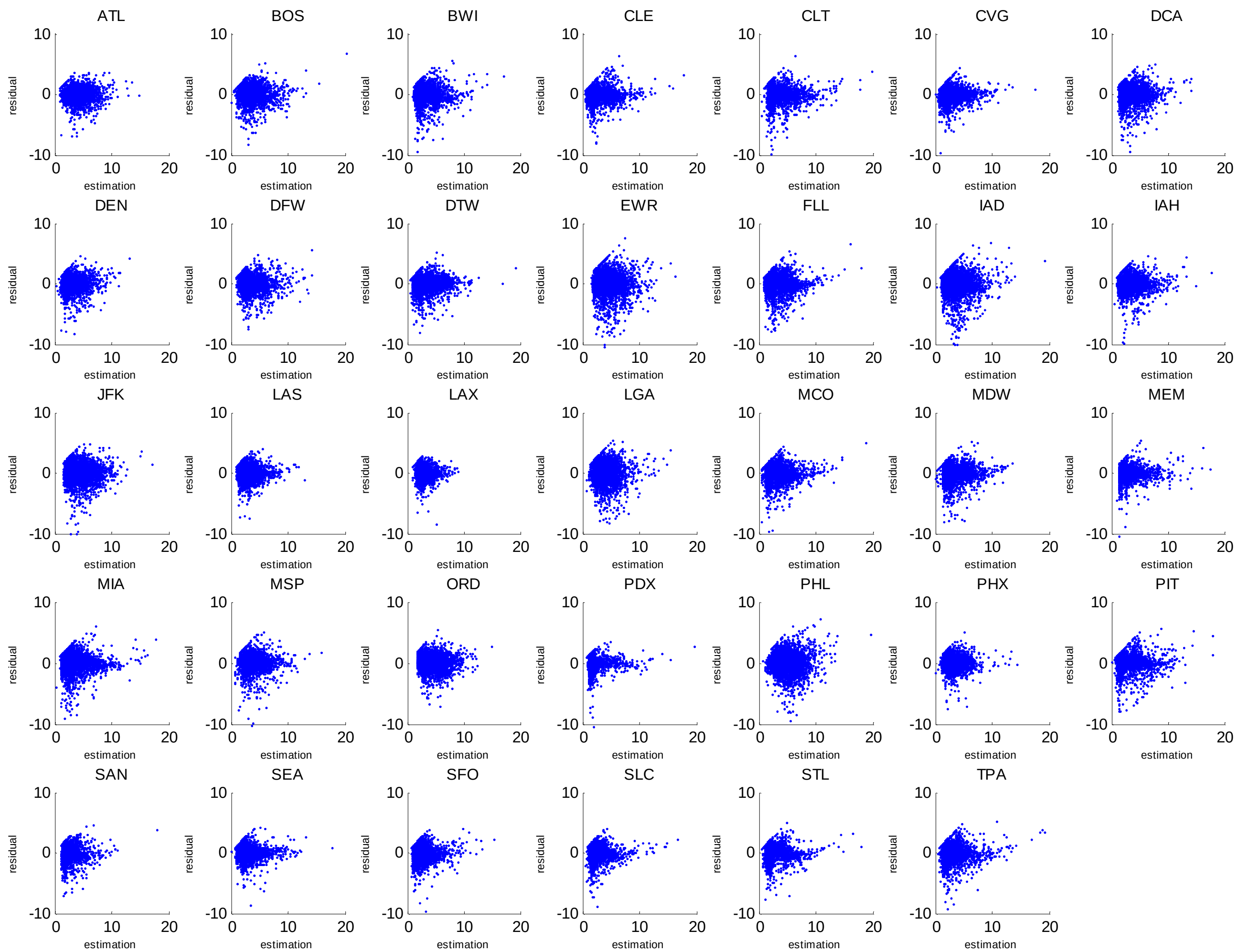


Figure **A.3**: Scatter Plot of Residuals of Squared Root of Generated Delay from Reduced-size Model

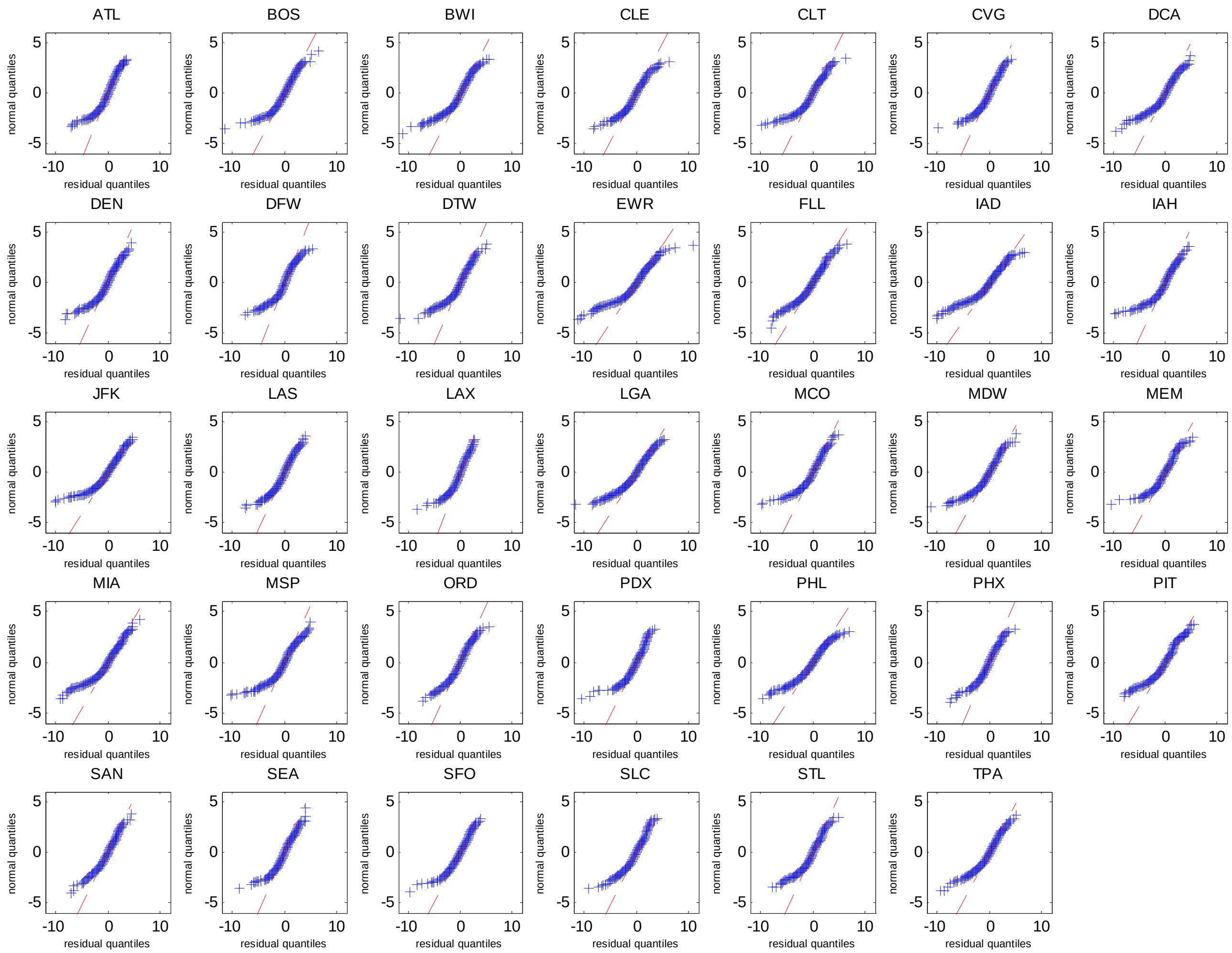


Figure **A.4**: Quantile_Quantile Plot of Residuals of Square Root of Generated Delay from Reduceced-size Model

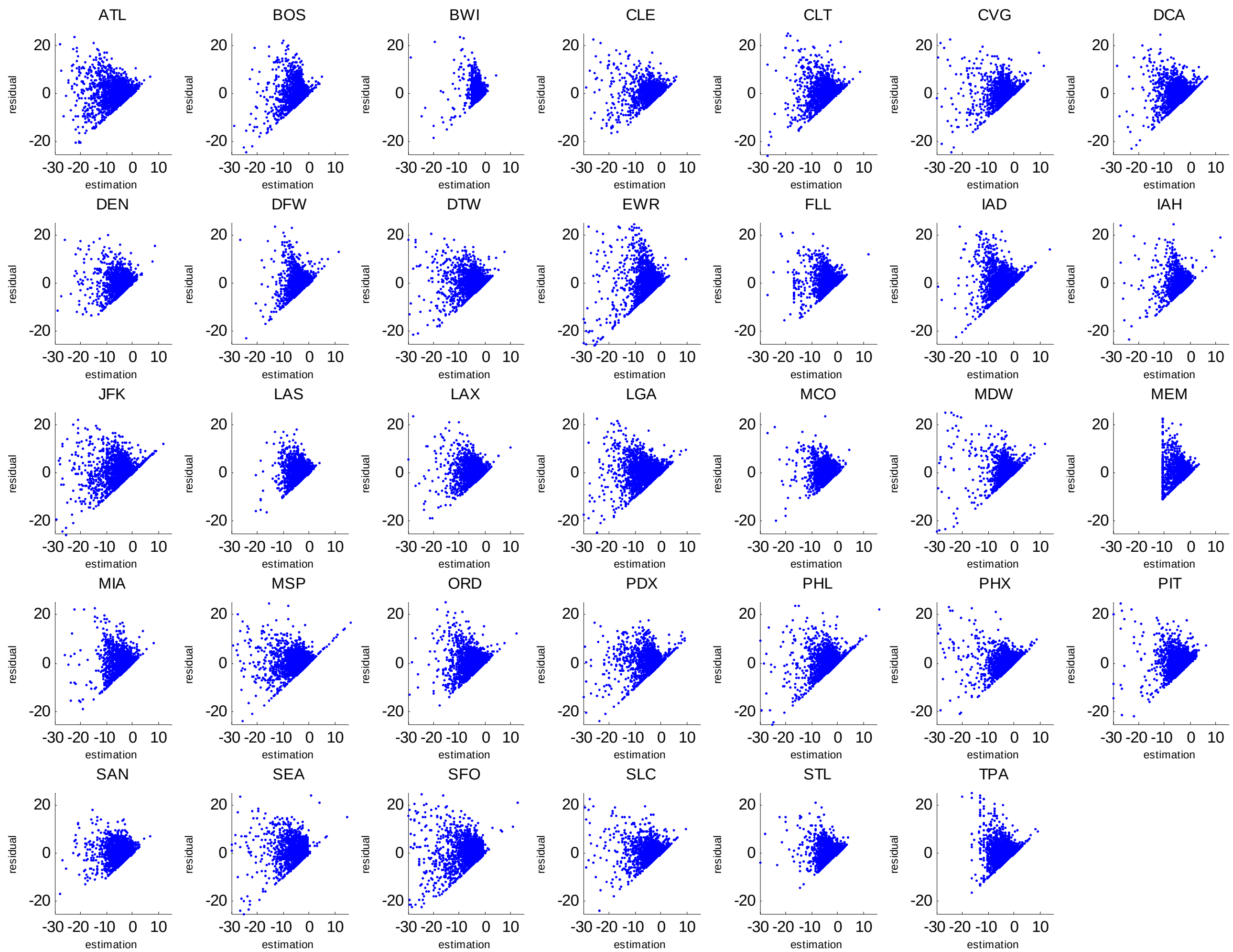


Figure **A.5**: Scatter Plot of Residuals of Absorbed Delay on Original Scale from Full-size Model

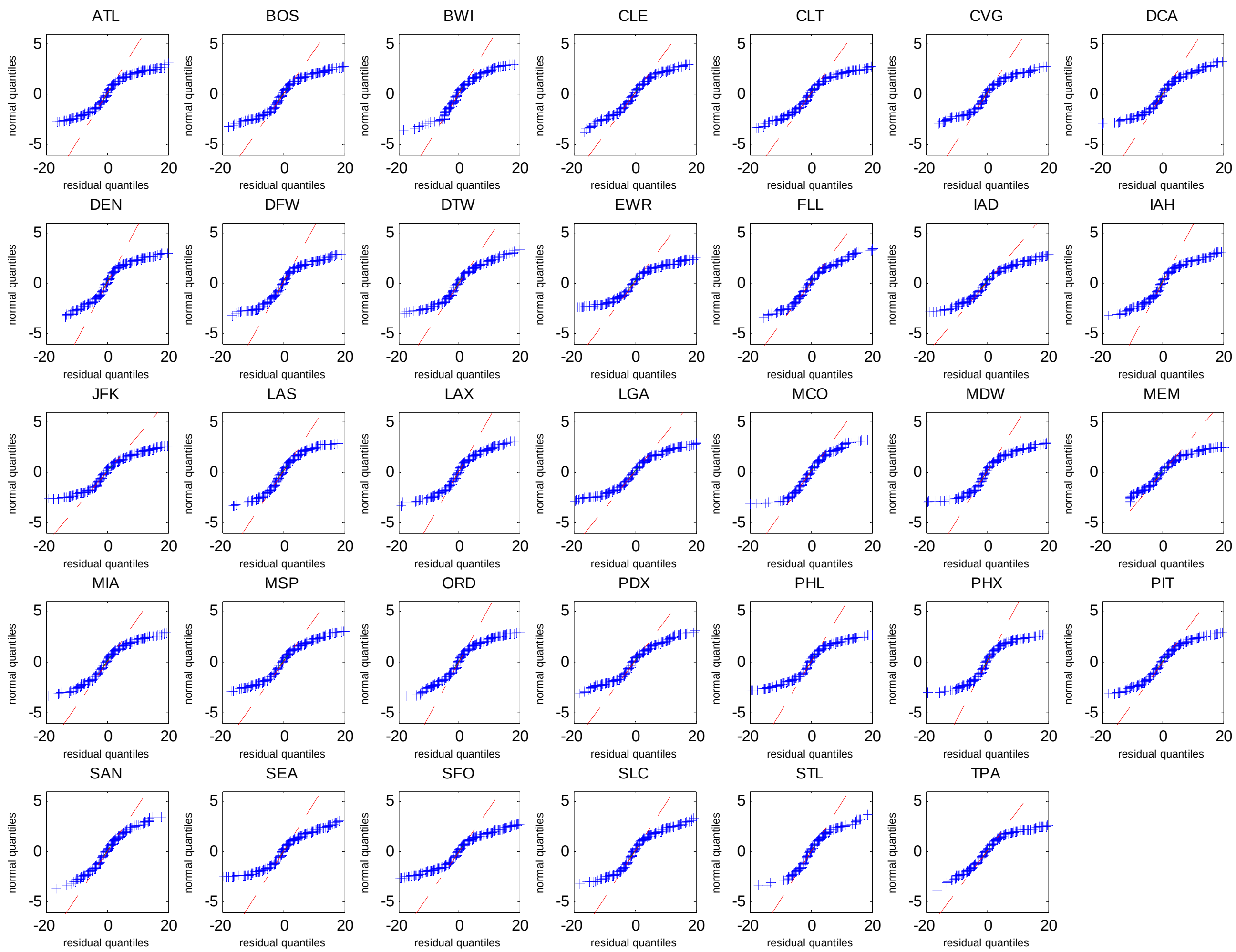


Figure **A.6**: Quantile-Quantile Plot of Residuals of Absorbed Delay on the Original Scale from Full-size Model

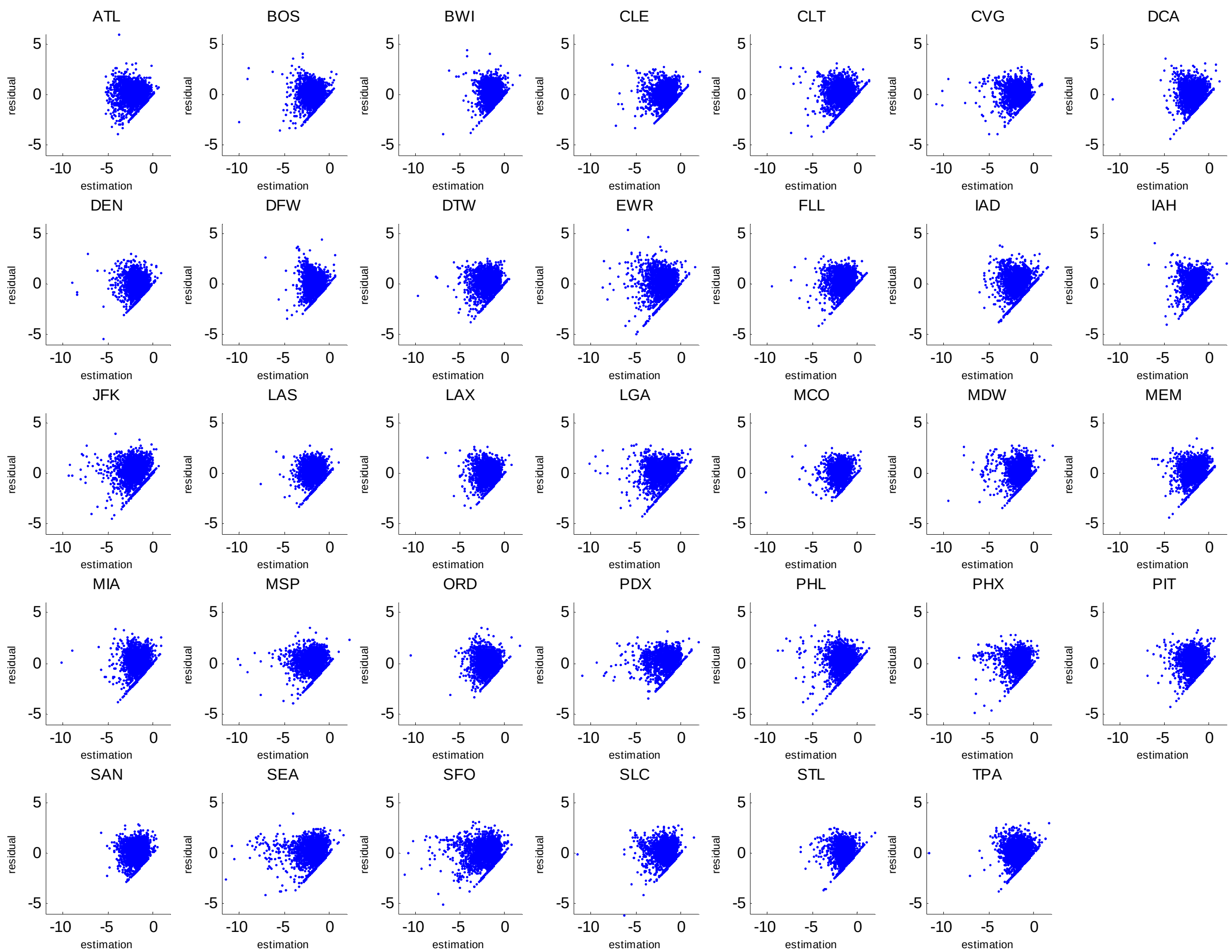


Figure **A.7**: Scatter Plot of Residuals of Squared Root of Absorbed Delay from Reduced-size Mode

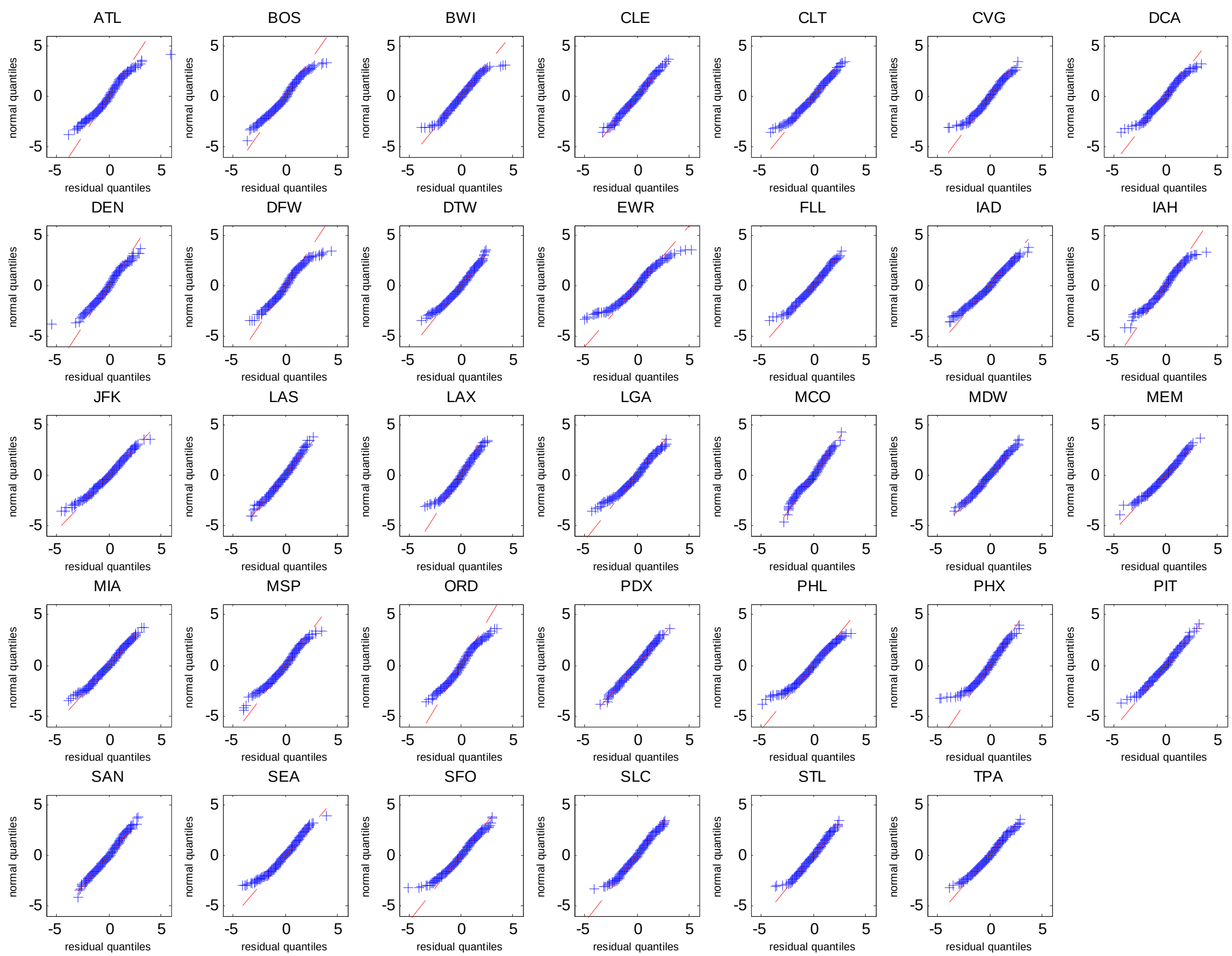


Figure **A.8**: Quantile-Quantile Plot of Residuals of Squared Root of Generated Delay from Reduced-size Mode

Table **A.3**: Test Result for Airport Generated Delay

Airport	Squared Residuals				Squared Prediction Errors			
	Paired t	p-value	ks	p-value	Paired t	p-value	ks	p-value
ATL	-0.19	0.4266	0.02	0.7229	1.43	0.9236	0.02	0.7229
BOS	-0.75	0.2280	0.02	0.7877	0.80	0.7874	0.02	0.7877
BWI	-0.65	0.2592	0.02	0.3703	-0.68	0.2487	0.02	0.3703
CLE	-1.46	0.0725	0.02	0.3842	0.97	0.8341	0.02	0.3842
CLT	2.34	0.9903	0.03	0.1480	0.32	0.6272	0.03	0.1480
CVG	-0.15	0.4385	0.01	0.9999	-0.23	0.4077	0.01	0.9999
DCA	0.57	0.7162	0.01	1.0000	1.38	0.9154	0.01	1.0000
DEN	-0.51	0.3063	0.00	1.0000	-0.52	0.3007	0.00	1.0000
DFW	-0.94	0.1747	0.01	0.8267	-1.10	0.1348	0.01	0.8267
DTW	0.30	0.6163	0.01	0.9852	0.42	0.6639	0.01	0.9852
EWR	-0.62	0.2674	0.03	0.0511	1.42	0.9213	0.03	0.0511
FLL	-0.53	0.2996	0.01	0.8776	-1.12	0.1322	0.01	0.8776
IAD	1.09	0.8627	0.02	0.3950	2.44	0.9925	0.02	0.3950
IAH	-0.58	0.2794	0.02	0.8424	1.53	0.9370	0.02	0.8424
JFK	-1.51	0.0660	0.01	0.9800	2.67	0.9961	0.01	0.9800
LAS	-0.72	0.2359	0.01	0.9556	2.77	0.9971	0.01	0.9556
LAX	-1.03	0.1515	0.02	0.7815	1.78	0.9619	0.02	0.7815
LGA	-0.18	0.4288	0.01	0.9846	-0.19	0.4232	0.01	0.9846
MCO	-0.87	0.1916	0.02	0.6811	0.22	0.5865	0.02	0.6811
MDW	-0.16	0.4348	0.02	0.3676	-0.59	0.2776	0.02	0.3676
MEM	0.01	0.5042	0.02	0.4553	1.11	0.8664	0.02	0.4553
MIA	-0.36	0.3579	0.01	0.9143	-1.04	0.1485	0.01	0.9143
MSP	-1.57	0.0580	0.01	0.9277	-0.30	0.3809	0.01	0.9277
ORD	NaN	NaN	0.00	1.0000	NaN	NaN	0.00	1.0000
PDX	-0.78	0.2177	0.02	0.6038	-1.46	0.0728	0.02	0.6038
PHL	0.25	0.5983	0.00	1.0000	-0.22	0.4134	0.00	1.0000
PHX	-1.36	0.0877	0.03	0.1645	2.44	0.9926	0.03	0.1645
PIT	-0.84	0.1995	0.02	0.2769	1.00	0.8412	0.02	0.2769
SAN	-0.92	0.1791	0.01	0.8110	0.55	0.7088	0.01	0.8110
SEA	0.74	0.7716	0.01	0.9430	1.16	0.8776	0.01	0.9430
SFO	1.44	0.9252	0.01	1.0000	-0.75	0.2275	0.01	1.0000
SLC	-1.03	0.1509	0.02	0.3934	-1.04	0.1504	0.02	0.3934
STL	-0.90	0.1852	0.02	0.5978	-0.28	0.3893	0.02	0.5978
TPA	-0.44	0.3316	0.01	0.9578	3.12	0.9991	0.01	0.9578

* P-vlaue greater than 0.05 means the null hypothesis cannot be rejected at 5% significant level.

Table **A.4**: Test Result for Airport Absorbed Delay

Airport	Squared Residuals				Squared Prediction Errors			
	Paired t	p-value	ks	p-value	Paired t	p-value	ks	p-value
ATL	1.76	0.9606	0.02	0.7416	-0.15	0.4407	0.02	0.7416
BOS	-0.58	0.2826	0.00	1.0000	-0.53	0.2983	0.00	1.0000
BWI	3.77	0.9999	0.02	0.5457	-1.02	0.1539	0.02	0.5457
CLE	-1.31	0.0946	0.02	0.7552	1.84	0.9673	0.02	0.7552
CLT	-0.21	0.4153	0.02	0.6372	-0.76	0.2227	0.02	0.6372
CVG	-1.53	0.0625	0.03	0.1761	1.74	0.9593	0.03	0.1761
DCA	-0.36	0.3589	0.01	0.9987	-0.76	0.2247	0.01	0.9987
DEN	-0.52	0.3020	0.01	0.9177	1.21	0.8859	0.01	0.9177
DFW	-1.23	0.1101	0.01	0.9973	-0.12	0.4511	0.01	0.9973
DTW	-1.50	0.0666	0.01	0.9891	-1.26	0.1042	0.01	0.9891
EWR	-0.97	0.1665	0.02	0.2658	-0.22	0.4119	0.02	0.2658
FLL	-0.52	0.3028	0.02	0.6588	1.60	0.9451	0.02	0.6588
IAD	-0.01	0.4978	0.00	1.0000	-1.49	0.0688	0.00	1.0000
IAH	-1.03	0.1516	0.02	0.8768	-0.09	0.4626	0.02	0.8768
JFK	-1.34	0.0896	0.02	0.3248	-0.30	0.3823	0.02	0.3248
LAS	1.49	0.9313	0.01	0.8280	1.80	0.9638	0.01	0.8280
LAX	-0.24	0.4061	0.00	1.0000	-0.74	0.2301	0.00	1.0000
LGA	-1.62	0.0526	0.01	0.9425	0.49	0.6875	0.01	0.9425
MCO	0.34	0.6320	0.02	0.7399	0.03	0.5125	0.02	0.7399
MDW	-0.75	0.2256	0.02	0.4351	-0.12	0.4511	0.02	0.4351
MEM	-0.42	0.3372	0.03	0.2863	0.31	0.6205	0.03	0.2863
MIA	-0.08	0.4668	0.03	0.2051	-0.24	0.4034	0.03	0.2051
MSP	-0.58	0.2804	0.02	0.5969	-0.10	0.4587	0.02	0.5969
ORD	0.04	0.5160	0.02	0.6824	-1.51	0.0656	0.02	0.6824
PDX	NaN	NaN	0.00	1.0000	NaN	NaN	0.00	1.0000
PHL	-0.92	0.1796	0.00	1.0000	-1.16	0.1224	0.00	1.0000
PHX	-0.54	0.2961	0.00	1.0000	0.89	0.8120	0.00	1.0000
PIT	0.15	0.5609	0.01	1.0000	0.32	0.6242	0.01	1.0000
SAN	0.46	0.6765	0.02	0.6408	1.34	0.9100	0.02	0.6408
SEA	0.48	0.6859	0.02	0.2611	1.41	0.9205	0.02	0.2611
SFO	-0.70	0.2407	0.01	0.9660	-0.33	0.3723	0.01	0.9660
SLC	-0.61	0.2710	0.01	0.9416	-0.31	0.3796	0.01	0.9416
STL	0.63	0.7348	0.01	1.0000	-0.37	0.3568	0.01	1.0000
TPA	0.37	0.6459	0.00	1.0000	1.73	0.9580	0.00	1.0000

* P-vlaue greater than 0.05 means the null hypothesis cannot be rejected at 5% significant level.

Table **A.5**: Predictors for Airport Generated Delay in the Reduced-size Models of 34 OEP Airports

	Predictor	ATL	BOS	BWI	CLE	CLT	CVG	DCA	DEN	DFW	DTW	EWR	FLL
Inbound	inbound delay	x	x	x	x	x	x	x	x	x	x		x
Airline	carrier delay	x	x	x	x	x	x	x	x	x	x	x	x
	swap aircraft rate	x	x	x	x	x	x	x	x	x		x	x
	cancelled departure rate									x			
	scheduled departure time	x	x	x	x		x			x	x	x	x
	leg number									x			
	scheduled turnaround time												
Departure Demand Ratio	departure rho (demand/throughput in 30min)	x		x	x			x	x	x		x	x
	departure rho (demand/throughput in 15min)												
	departure rho (demand/ADR in 30min)		x				x						
	departure rho (demand/ADR in 15min)					x		x			x		
	ADR												
	arrival rho (demand/throughput in 30min)												
	arrival rho (demand/AAR in 30min)							x		x			
	arrival rho (demand/AAR in 15min)								x				
	AAR								x				
Airport	terminal weather			x		x		x		x			
	visibility								x				
	runway configuration												
	IMC / VMC												
	security delay												
Outbound	GDP holding time	x	x	x	x	x	x	x	x	x	x	x	x
	actual enroute time weather						x	x					
	scheduled enroute time weather		x										
	outbound distance												

	Predictor	IAD	IAH	JFK	LAS	LAX	LGA	MCO	MDW	MEM	MIA	MSP	ORD
Inbound	inbound delay	x	x	x	x	x	x	x	x	x	x	x	x
Airline	carrier delay	x	x	x	x	x	x	x	x	x	x	x	x
	swap aircraft rate	x		x	x	x	x	x	x	x			x
	cancelled departure rate												
	scheduled departure time	x	x	x	x	x	x	x	x		x	x	x
	leg number												
	scheduled turnaround time								x				
Departure Demand Ratio	departure rho (demand/throughput in 30min)	x		x	x	x	x	x			x		x
	departure rho (demand/throughput in 15min)												
	departure rho (demand/ADR in 30min)												
	departure rho (demand/ADR in 15min)		x							x		x	
	ADR		x										
	arrival rho (demand/throughput in 30min)												
	arrival rho (demand/AAR in 30min)												
	arrival rho (demand/AAR in 15min)												
	AAR		x		x								
Airport	terminal weather		x				x	x		x	x	x	x
	visibility												
	runway configuration												x
	IMC / VMC												
	security delay								x				
Outbound	GDP holding time	x	x	x	x	x	x	x	x	x	x	x	x
	actual enroute time weather						x						x
	scheduled enroute time weather								x				
	outbound distance												

	Predictor	PDX	PHL	PHX	PIT	SAN	SEA	SFO	SLC	STL	TPA
Inbound	inbound delay	x		x	x	x	x	x	x	x	x
Airline	carrier delay	x	x	x	x	x	x	x	x	x	x
	swap aircraft rate	x	x	x	x	x	x	x	x		x
	cancelled departure rate										
	scheduled departure time	x	x	x		x	x			x	x
	leg number							x			
	scheduled turnaround time	x									
Departure Demand Ratio	departure rho (demand/throughput in 30min)	x	x	x	x		x			x	x
	departure rho (demand/throughput in 15min)					x					
	departure rho (demand/ADR in 30min)										
	departure rho (demand/ADR in 15min)							x	x		
	ADR										
	arrival rho (demand/throughput in 30min)		x								
	arrival rho (demand/AAR in 30min)								x		
	arrival rho (demand/AAR in 15min)										
	AAR			x							
Airport	terminal weather				x						
	visibility					x					
	runway configuration		x								
	IMC / VMC					x					
	security delay										
Outbound	GDP holding time	x	x	x	x	x	x	x	x	x	x
	actual enroute time weather		x								
	scheduled enroute time weather		x								
	outbound distance						x	x			

Table **A.6**: Predictors for Airport Absorbed Delay in the Reduced-size Models of 34 OEP Airports

	factor	ATL	BOS	BWI	CLE	CLT	CVG	DCA	DEN	DFW	DTW	EWR	FLL
Inbound	inbound delay	x	x	x	x	x	x	x	x	x	x	x	x
Airline	carrier delay	x	x	x	x	x	x	x	x	x	x	x	x
	scheduled turnaround time	x	x	x	x	x	x	x	x	x	x	x	x
	SEATS	x		x		x	x	x	x	x		x	x
	WEIGHT				x								
	scheduled departure time				x								x
	leg number												
Departure Demand Ratio	departure rho (demand/throughput in 30min)											x	
	departure rho (demand/ADR both in 15min)												
	departure rho (demand/ADR in 15min)												
Outbound	GDP holding time	x	x	x	x	x	x	x	x	x	x	x	x
	scheduled enroute time weather												
	destination arrival rho (demand/AAR in 30min)				x								
	outbound distiance		x	x									

	Predictor	IAD	IAH	JFK	LAS	LAX	LGA	MCO	MDW	MEM	MIA	MSP	ORD
Inbound	inbound delay	x	x	x	x	x	x	x	x	x	x	x	x
Airline	carrier delay	x	x	x	x	x	x	x	x	x	x	x	x
	scheduled turnaround time	x	x	x	x	x	x	x	x	x	x	x	x
	SEATS	x	x		x	x	x	x		x	x		x
	WEIGHT												
	scheduled departure time				x								
	leg number	x											
Departure Demand Ratio	departure rho (demand/throughput in 30min)	x											
	departure rho (demand/ADR both in 15min)					x							
	departure rho (demand/ADR in 15min)					x							
Outbound	GDP holding time	x	x				x	x	x	x	x	x	x
	scheduled enroute time weather												
	destination arrival rho (demand/AAR in 30min)												
	outbound distiance					x	x						

	Predictor	PDX	PHL	PHX	PIT	SAN	SEA	SFO	SLC	STL	TPA
Inbound	inbound delay	x	x	x	x	x	x	x	x	x	x
Airline	carrier delay	x	x	x	x	x	x	x	x	x	x
	scheduled turnaround time	x	x	x	x	x	x	x	x	x	x
	SEATS	x	x		x	x	x			x	
	WEIGHT		x						x		
	scheduled departure time		x					x		x	x
	leg number	x		x							
Departure Demand Ratio	departure rho (demand/throughput in 30min)										
	departure rho (demand/ADR both in 15min)					x					
	departure rho (demand/ADR in 15min)										
Outbound	GDP holding time	x	x		x		x	x	x	x	x
	scheduled enroute time weather				x						
	destination arrival rho (demand/AAR in 30min)										
	outbound distiance				x						x

Table **A.7**: Percentage of Actual Airport Delay Data (w) in Validation Set of 2005 and 2006 in Regression Value $\pm\hat{\sigma}$ and $2\hat{\sigma}$.

Airport Code	Percentage of actual data in 2005		Percentage of actual data in 2006	
	Within 68%	Within 95%	Within 68%	Within 95%
ATL	75.9%	94.9%	72.6%	92.7%
BOS	76.1%	94.2%	74.3%	94.9%
BWI	75.3%	95.1%	79.0%	96.8%
CLE	74.6%	91.1%	70.4%	91.6%
CLT	77.2%	95.9%	71.1%	93.8%
CVG	67.9%	92.7%	67.0%	92.4%
DCA	83.6%	97.3%	77.9%	95.0%
DEN	75.6%	94.5%	73.0%	95.8%
DFW	77.7%	94.3%	78.0%	93.2%
DTW	79.5%	95.3%	77.1%	95.8%
EWR	86.1%	98.2%	73.6%	94.2%
FLL	82.2%	96.6%	71.1%	95.7%
IAD	84.3%	96.1%	80.1%	96.1%
IAH	73.5%	90.6%	71.2%	91.9%
JFK	79.8%	96.5%	72.5%	90.7%
LAS	73.3%	95.1%	68.2%	91.2%
LAX	75.6%	96.2%	76.1%	93.8%
LGA	78.4%	95.1%	73.6%	95.1%
MCO	78.0%	95.4%	74.0%	93.4%
MDW	79.0%	95.9%	76.1%	94.9%
MEM	70.1%	91.0%	67.8%	92.8%
MIA	78.8%	94.9%	72.6%	91.6%
MSP	83.0%	97.1%	74.6%	96.0%
ORD	81.6%	96.9%	72.5%	94.5%
PDX	72.7%	94.1%	75.7%	94.1%
PHL	79.0%	97.6%	71.1%	95.3%
PHX	71.5%	94.5%	67.4%	90.7%
PIT	77.2%	94.5%	73.3%	92.4%
SAN	73.9%	91.6%	69.9%	93.5%
SEA	78.9%	97.0%	77.1%	93.7%
SFO	72.4%	93.4%	72.8%	94.1%
SLC	76.4%	93.6%	71.0%	90.2%
STL	73.9%	94.3%	73.9%	94.8%
TPA	75.7%	94.3%	71.5%	90.5%
Average	**77.0%**	**94.9%**	**73.2%**	**93.6%**
min	**67.9%**	**90.6%**	**67.0%**	**90.2%**
max	**86.1%**	**98.2%**	**80.1%**	**96.8%**

APPENDIX B

FINAL MODELS

B.1 Airport Generated Delay Models

Airport: ATL

Basis Functions
===============

BF1 = max(0, GDPHoldingTime - 23.200);
BF2 = max(0, 23.200 - GDPHoldingTime);
BF3 = max(0, CarrierDelay - 11.857) * BF2;
BF4 = max(0, 11.857 - CarrierDelay) * BF2;
BF6 = max(0, 34.000 - ScheduleDepartureTime) * BF2;
BF7 = max(0, DepartureDemandRatio30 - 6.160);
BF8 = max(0, 6.160 - DepartureDemandRatio30);
BF9 = max(0, SwapAircraftRate - 0.330);
BF10 = max(0, 0.330 - SwapAircraftRate);
BF12 = max(0, 56.000 - CarrierDelay) * BF10;
BF13 = max(0, InboundDelay - 124.250) * BF10;
BF14 = max(0, 124.250 - InboundDelay) * BF10;
BF15 = max(0, ScheduleDepartureTime - 54.000);
BF16 = max(0, 54.000 - ScheduleDepartureTime);
BF17 = max(0, CarrierDelay - 3.333) * BF16;
BF18 = max(0, 3.333 - CarrierDelay) * BF16;

sqrt(GeneratedDelay) = max(0, 10.785 + 0.041 * BF1 - 0.074 * BF2 + .665814E-03 * BF3
- 0.003 * BF4 - 0.010 * BF6 - 0.030 * BF7 - 0.653 * BF8
+ 1.934 * BF9 - 1.150 * BF10 - 0.163 * BF12
- 0.092 * BF13 + 0.031 * BF14 + 0.027 * BF15
+ 0.033 * BF16 + .961083E-03 * BF17 - 0.005 * BF18)

Airport: BOS

Basis Functions
===============

BF1 = max(0, GDPHoldingTime - 6.200);
BF2 = max(0, 6.200 - GDPHoldingTime);
BF3 = max(0, CarrierDelay - 3.167);
BF4 = max(0, 3.167 - CarrierDelay);

BF5 = max(0, InboundDelay - 139.000) * BF1;
BF6 = max(0, 139.000 - InboundDelay) * BF1;
BF7 = max(0, RD30SS - 0.542);
BF8 = max(0, 0.542 - RD30SS);
BF9 = max(0, SwapAircraftRate + .707592E-09) * BF2;
BF10 = max(0, ScheduleEnrouteWeather - 2054.250);
BF11 = max(0, 2054.250 - ScheduleEnrouteWeather);
BF12 = max(0, ScheduleDepartureTime - 55.000) * BF10;
BF13 = max(0, 55.000 - ScheduleDepartureTime) * BF10;
BF14 = max(0, ScheduleDepartureTime - 48.000) * BF7;
BF15 = max(0, 48.000 - ScheduleDepartureTime) * BF7;
BF17 = max(0, 8.333 - CarrierDelay) * BF2;
BF18 = max(0, SwapAircraftRate + .707592E-09) * BF4;

sqrt(GeneratedDelay) = max(0, 4.159 - 0.024 * BF1 - 0.024 * BF2 + 0.066 * BF3 - 0.080 * BF4
+ .486045E-03 * BF5 + .721596E-03 * BF6 + 0.297 * BF7
- 3.198 * BF8 + 0.648 * BF9 + .294995E-03 * BF10
- .265383E-03 * BF11 - .121340E-04 * BF12 - .128603E-04 * BF13
+ 0.083 * BF14 + 0.045 * BF15 - 0.020 * BF17
+ 1.154 * BF18)

Airport: BWI

Basis Functions
===============

BF1 = max(0, GDPHoldingTime - 15.500);
BF2 = max(0, 15.500 - GDPHoldingTime);
BF3 = max(0, CarrierDelay - 5.000) * BF2;
BF4 = max(0, 5.000 - CarrierDelay) * BF2;
BF5 = max(0, DepartureDemandRatio30 - 3.000) * BF2;
BF6 = max(0, 3.000 - DepartureDemandRatio30) * BF2;
BF7 = max(0, SwapAircraftRate + .719125E-10);
BF8 = max(0, InboundDelay + 28.500) * BF1;
BF9 = max(0, CarrierDelay - 6.000) * BF7;
BF10 = max(0, 6.000 - CarrierDelay) * BF7;
BF11 = max(0, ScheduleDepartureTime - 73.000);
BF12 = max(0, 73.000 - ScheduleDepartureTime);
BF13 = (TerminalWeather = 8 OR TerminalWeather = 9) * BF12;
BF15 = max(0, ScheduleDepartureTime - 30.000) * BF2;
BF16 = max(0, 30.000 - ScheduleDepartureTime) * BF2;
BF17 = max(0, ScheduleDepartureTime - 51.000);

sqrt(GeneratedDelay) = max(0, 3.520 + 0.081 * BF1 - 0.035 * BF2 + 0.005 * BF3 - 0.017 * BF4
+ 0.017 * BF5 - 0.048 * BF6 + 4.024 * BF7 - .516493E-03 * BF8
- 0.067 * BF9 + 0.713 * BF10 - 0.152 * BF11
+ 0.046 * BF12 + 0.158 * BF13 + 0.002 * BF15
- 0.014 * BF16 + 0.055 * BF17)

Airport: CLE

Basis Functions
===============
BF1 = max(0, GDPHoldingTime - 14.800);
BF2 = max(0, 14.800 - GDPHoldingTime);
BF3 = max(0, CarrierDelay - 0.333) * BF2;
BF4 = max(0, 0.333 - CarrierDelay) * BF2;
BF5 = max(0, InboundDelay - 77.885) * BF1;
BF6 = max(0, 77.885 - InboundDelay) * BF1;
BF7 = max(0, DepartureDemandRatio30 - 3.050);
BF8 = max(0, 3.050 - DepartureDemandRatio30);
BF10 = max(0, 0.250 - SwapAircraftRate);
BF11 = max(0, ScheduleDepartureTime - 76.000) * BF10;
BF14 = max(0, 10.167 - CarrierDelay) * BF8;
BF16 = max(0, 31.000 - ScheduleDepartureTime) * BF8;
BF17 = max(0, ScheduleDepartureTime - 45.000) * BF2;
BF18 = max(0, 45.000 - ScheduleDepartureTime) * BF2;

sqrt(GeneratedDelay) = max(0, 6.499 + 0.013 * BF1 - 0.151 * BF2 + 0.004 * BF3 - 0.115 * BF4
- .287213E-03 * BF5 + .788110E-03 * BF6 - 0.085 * BF7
- 6.096 * BF10 - 0.636 * BF11 - 0.034 * BF14
- 0.125 * BF16 + 0.002 * BF17 + 0.003 * BF18)

Airport: CLT

Basis Functions
===============

BF1 = max(0, GDPHoldingTime - 9.100);
BF2 = max(0, 9.100 - GDPHoldingTime);
BF3 = max(0, CarrierDelay - 0.400) * BF2;
BF4 = max(0, 0.400 - CarrierDelay) * BF2;
BF5 = max(0, InboundDelay - 109.000) * BF1;
BF6 = max(0, 109.000 - InboundDelay) * BF1;
BF7 = max(0, SwapAircraftRate - 0.200);
BF8 = max(0, 0.200 - SwapAircraftRate);
BF10 = max(0, 0.429 - DepartureDemandRatio_ADR15);
BF11 = max(0, CarrierDelay - 43.000);
BF12 = max(0, 43.000 - CarrierDelay);
BF13 = (TerminalWeather = 0 OR TerminalWeather = 1 OR TerminalWeather = 2 OR TerminalWeather = 3
OR TerminalWeather = 4 OR TerminalWeather = 5 OR TerminalWeather = 7) * BF2;
BF15 = max(0, InboundDelay + 7.667);
BF16 = max(0, - 7.667 - InboundDelay);
BF18 = max(0, 48.000 - CarrierDelay) * BF15;

sqrt(GeneratedDelay) = max(0, 9.152 + 0.119 * BF2 + 0.006 * BF3 - 0.164 * BF4 + .580519E-03 * BF5
+ .636105E-03 * BF6 + 2.909 * BF7 - 8.790 * BF8
- 3.262 * BF10 - 0.022 * BF11 - 0.069 * BF12
- 0.235 * BF13 - 0.046 * BF15 - 0.030 * BF16
+ .830188E-03 * BF18)

Airport: CVG

```
Basis Functions
===============
BF1 = max(0, GDPHoldingTime - 15.800);
BF2 = max(0, 15.800 - GDPHoldingTime );
BF3 = max(0, CarrierDelay - 4.429) * BF2;
BF4 = max(0, 4.429 - CarrierDelay ) * BF2;
BF6 = max(0, 0.656 - RD30SS );
BF7 = max(0, InboundDelay - 47.000) * BF1;
BF8 = max(0, 47.000 - InboundDelay ) * BF1;
BF9 = max(0, SwapAircraftRate - .357727E-09);
BF10 = max(0, ActuralEnrouteWeather + .586138E-04);
BF11 = max(0, InboundDelay + 38.000);
BF12 = max(0, GDPHoldingTime - 109.000) * BF11;
BF13 = max(0, 109.000 - GDPHoldingTime ) * BF11;
BF15 = max(0, 55.000 - CarrierDelay );
BF16 = max(0, ScheduleDepartureTime - 42.000) * BF6;
BF18 = max(0, InboundDelay + 38.000) * BF15;

sqrt(GeneratedDelay) = max(0, 10.853 + 0.175 * BF1 - 0.164 * BF2 + 0.002 * BF3 - 0.011 * BF4
        - 2.662 * BF6 + 0.001 * BF7 - .956928E-03 * BF8
        + 2.798 * BF9 + .141702E-03 * BF10 - 0.218 * BF11
        - 0.001 * BF12 + 0.002 * BF13 - 0.087 * BF15
        + 0.051 * BF16 + .778851E-03 * BF18)
```

Airport: DCA

```
Basis Functions
===============

BF1 = max(0, GDPHoldingTime - 13.600);
BF2 = max(0, 13.600 - GDPHoldingTime );
BF3 = max(0, CarrierDelay - 1.500) * BF2;
BF4 = max(0, 1.500 - CarrierDelay ) * BF2;
BF6 = max(0, 7229.000 - ActuralEnrouteWeather );
BF7 = max(0, SwapAircraftRate + .138945E-09);
BF9 = max(0, 50.750 - InboundDelay ) * BF1;
BF11 = max(0, 6.080 - DepartureDemandRatio30 );
BF12 = max(0, ArrivalDemandRatio_AAR30 - 0.600) * BF2;
BF15 = max(0, 39.667 - CarrierDelay ) * BF11;
BF16 = ( TerminalWeather = 8);
BF18 = max(0, DepartureDemandRatio_ADR15 - 0.100) * BF2;

sqrt(GeneratedDelay) = max(0, 7.929 + 0.023 * BF1 - 0.129 * BF2 + 0.004 * BF3 - 0.032 * BF4
        - .247570E-03 * BF6 + 5.325 * BF7 + 0.001 * BF9
        + 0.101 * BF12 - 0.011 * BF15 + 1.982 * BF16
        + 0.031 * BF18)
```

Airport: DEN

Basis Functions
===============

BF1 = max(0, CarrierDelay - 0.167);
BF2 = max(0, 0.167 - CarrierDelay);
BF3 = max(0, SwapAircraftRate - 0.250);
BF4 = max(0, 0.250 - SwapAircraftRate);
BF5 = max(0, GDPHoldingTime - 14.000);
BF6 = max(0, 14.000 - GDPHoldingTime);
BF7 = max(0, DepartureDemandRatio30 - 8.730);
BF8 = max(0, 8.730 - DepartureDemandRatio30);
BF9 = max(0, ArrivalDemandRatio_AAR15 - 0.222) * BF4;
BF10 = max(0, 0.222 - ArrivalDemandRatio_AAR15) * BF4;
BF11 = max(0, AAR - 24.000) * BF4;
BF12 = max(0, 24.000 - AAR) * BF4;
BF13 = max(0, VISIB_CR - 8.000);
BF14 = max(0, 8.000 - VISIB_CR);
BF15 = max(0, CarrierDelay - 20.500) * BF6;
BF16 = max(0, 20.500 - CarrierDelay) * BF6;
BF17 = max(0, InboundDelay + 21.500) * BF13;
BF18 = max(0, - 21.500 - InboundDelay) * BF13;

sqrt(GeneratedDelay) = max(0, 12.043 + 0.029 * BF1 - 2.511 * BF2 + 4.686 * BF3 - 8.155 * BF4
+ 0.060 * BF5 + 0.025 * BF6 + 0.044 * BF7 - 0.459 * BF8
- 1.237 * BF9 - 17.500 * BF10 - 0.080 * BF11
+ 0.942 * BF12 - 0.861 * BF13 - 0.365 * BF14
+ 0.003 * BF15 - 0.007 * BF16 - 0.006 * BF17
- 0.060 * BF18)

Airport: DFW

Basis Functions
===============

BF1 = max(0, GDPHoldingTime - 16.200);
BF2 = max(0, 16.200 - GDPHoldingTime);
BF3 = max(0, CarrierDelay - 10.143) * BF2;
BF4 = max(0, 10.143 - CarrierDelay) * BF2;
BF5 = (TerminalWeather = 2 OR TerminalWeather = 8 OR TerminalWeather = 9);
BF7 = max(0, ScheduleDepartureTime - 39.000);
BF8 = max(0, 39.000 - ScheduleDepartureTime);
BF10 = max(0, 3.970 - DepartureDemandRatio30);
BF11 = max(0, InboundDelay + 26.000) * BF1;
BF12 = max(0, ArrivalDemandRatio_AAR30 - 0.821);
BF13 = max(0, 0.821 - ArrivalDemandRatio_AAR30);
BF14 = max(0, SwapAircraftRate + .471008E-09) * BF10;
BF15 = max(0, LegNumber - 1.000);
BF16 = max(0, CancelledDepartureRate - 0.036);
BF18 = (TerminalWeather = 4 OR TerminalWeather = 8 OR TerminalWeather = 9) * BF16;

sqrt(GeneratedDelay) = max(0, 7.059 + 0.078 * BF1 - 0.044 * BF2 + 0.005 * BF3 - 0.011 * BF4

+ 2.451 * BF5 + 0.032 * BF7 - 0.088 * BF8 - 0.579 * BF10
- .625545E-03 * BF11 + 28.937 * BF12 - 0.554 * BF13
+ 1.730 * BF14 - 0.391 * BF15 + 5.362 * BF16
+ 24.259 * BF18)

Airport: DTW

Basis Functions
===============

BF1 = max(0, CarrierDelay - 0.625);
BF2 = max(0, 0.625 - CarrierDelay);
BF3 = max(0, GDPHoldingTime - 12.000) * BF2;
BF4 = max(0, 12.000 - GDPHoldingTime) * BF2;
BF5 = max(0, DepartureDemandRatio_ADR15 - 0.196);
BF6 = max(0, 0.196 - DepartureDemandRatio_ADR15);
BF7 = max(0, GDPHoldingTime - 54.500);
BF8 = max(0, 54.500 - GDPHoldingTime);
BF9 = max(0, InboundDelay - 114.000) * BF7;
BF10 = max(0, 114.000 - InboundDelay) * BF7;
BF11 = max(0, CarrierDelay - 37.000) * BF8;
BF12 = max(0, 37.000 - CarrierDelay) * BF8;
BF13 = max(0, ScheduleDepartureTime - 63.000);
BF14 = max(0, 63.000 - ScheduleDepartureTime);
BF15 = max(0, InboundDelay - 121.000);
BF16 = max(0, 121.000 - InboundDelay);
BF17 = max(0, GDPHoldingTime - 70.600) * BF16;
BF18 = max(0, 70.600 - GDPHoldingTime) * BF16;

sqrt(GeneratedDelay) = max(0, 1.761 + 0.006 * BF1 + 1.145 * BF2 - 0.022 * BF3 - 0.171 * BF4
+ 0.683 * BF5 - 6.338 * BF6 + 0.034 * BF7 + 0.086 * BF8
- .277391E-03 * BF9 - .224443E-03 * BF10 + .812947E-03 * BF11
- 0.002 * BF12 - 0.024 * BF13 - 0.024 * BF14
+ 0.012 * BF15 + 0.060 * BF16 + .344492E-03 * BF17
- .788747E-03 * BF18)

Airport: EWR

Basis Functions
===============

BF1 = max(0, CarrierDelay - 14.857);
BF2 = max(0, 14.857 - CarrierDelay);
BF3 = max(0, GDPHoldingTime - 15.600) * BF2;
BF4 = max(0, 15.600 - GDPHoldingTime) * BF2;
BF6 = max(0, 5.110 - DepartureDemandRatio30);
BF7 = max(0, ScheduleDepartureTime - 29.000);
BF8 = max(0, 29.000 - ScheduleDepartureTime);
BF9 = max(0, ScheduleDepartureTime - 46.000) * BF6;
BF11 = max(0, ScheduleDepartureTime - 73.000);
BF12 = max(0, 73.000 - ScheduleDepartureTime);
BF14 = max(0, 52.400 - GDPHoldingTime) * BF12;
BF15 = max(0, ScheduleDepartureTime - 38.000) * BF6;

BF17 = max(0, SwapAircraftRate - 0.250) * BF2;
BF18 = max(0, 0.250 - SwapAircraftRate) * BF2;

sqrt(GeneratedDelay) = max(0, 15.749 + 0.029 * BF1 + 0.080 * BF2 + 0.002 * BF3 - 0.003 * BF4
- 0.496 * BF6 - 0.086 * BF7 - 0.203 * BF8 + 0.095 * BF9
- 0.083 * BF11 - 0.003 * BF14 - 0.092 * BF15
+ 0.278 * BF17 - 0.819 * BF18)

Airport: FLL

Basis Functions
===============

BF1 = max(0, GDPHoldingTime - 13.200);
BF2 = max(0, 13.200 - GDPHoldingTime);
BF3 = max(0, CarrierDelay - 2.000) * BF2;
BF4 = max(0, 2.000 - CarrierDelay) * BF2;
BF6 = max(0, 109.333 - InboundDelay) * BF1;
BF8 = max(0, 46.000 - ScheduleDepartureTime) * BF2;
BF9 = max(0, DepartureDemandRatio30 - 0.830);
BF12 = max(0, 0.500 - SwapAircraftRate) * BF2;
BF13 = max(0, ScheduleDepartureTime - 78.000);

sqrt(GeneratedDelay) = max(0, 4.112 + 0.240 * BF2 + 0.006 * BF3 - 0.046 * BF4 + .682817E-03 * BF6
- 0.006 * BF8 + 0.531 * BF9 - 0.542 * BF12 - 0.174 * BF13)

Airport: IAD

Basis Functions
===============

BF1 = max(0, CarrierDelay - 0.333);
BF2 = max(0, 0.333 - CarrierDelay);
BF3 = max(0, GDPHoldingTime - 27.000);
BF4 = max(0, 27.000 - GDPHoldingTime);
BF5 = max(0, SwapAircraftRate - 0.330) * BF4;
BF6 = max(0, 0.330 - SwapAircraftRate) * BF4;
BF8 = max(0, 72.000 - ScheduleDepartureTime) * BF2;
BF9 = max(0, InboundDelay - 105.000) * BF3;
BF10 = max(0, 105.000 - InboundDelay) * BF3;
BF11 = max(0, ScheduleDepartureTime - 47.000) * BF4;
BF12 = max(0, 47.000 - ScheduleDepartureTime) * BF4;
BF13 = max(0, ScheduleDepartureTime - 35.000) * BF4;
BF15 = max(0, DepartureDemandRatio30 - 6.870);
BF16 = max(0, 6.870 - DepartureDemandRatio30);
BF17 = max(0, ScheduleDepartureTime - 70.000);
BF18 = max(0, 70.000 - ScheduleDepartureTime);

sqrt(GeneratedDelay) = max(0, 8.410 + 0.067 * BF1 - 0.993 * BF2 - 0.013 * BF3 + 0.163 * BF4
+ 0.230 * BF5 - 0.412 * BF6 - 0.065 * BF8 - .885071E-04 * BF9
+ .752824E-03 * BF10 + 0.015 * BF11 - 0.004 * BF12
- 0.013 * BF13 + 0.125 * BF15 - 0.328 * BF16
- 0.138 * BF17 - 0.045 * BF18)

Airport: IAH

Basis Functions
===============

BF1 = max(0, GDPHoldingTime - 5.200);
BF2 = max(0, 5.200 - GDPHoldingTime);
BF3 = max(0, CarrierDelay - 0.333) * BF2;
BF4 = max(0, 0.333 - CarrierDelay) * BF2;
BF5 = max(0, DepartureDemandRatio_ADR15 - 0.786);
BF6 = max(0, 0.786 - DepartureDemandRatio_ADR15);
BF7 = (TerminalWeather = 8);
BF9 = max(0, ScheduleDepartureTime - 76.000);
BF10 = max(0, 76.000 - ScheduleDepartureTime);
BF11 = max(0, AAR - 20.000);
BF12 = max(0, 20.000 - AAR);
BF13 = max(0, ADR - 13.000);
BF14 = max(0, InboundDelay - 78.000);
BF15 = max(0, 78.000 - InboundDelay);
BF16 = max(0, CarrierDelay - 68.000) * BF15;
BF17 = max(0, 68.000 - CarrierDelay) * BF15;
BF18 = (TerminalWeather = 2 OR TerminalWeather = 8) * BF6;

sqrt(GeneratedDelay) = max(0, 5.234 + 0.056 * BF1 - 0.127 * BF2 + 0.008 * BF3 - 0.248 * BF4
+ 0.314 * BF5 - 1.557 * BF6 + 3.677 * BF7 - 0.139 * BF9
- 0.028 * BF10 - 0.094 * BF11 + 0.170 * BF12
- 0.254 * BF13 - 0.113 * BF14 + 0.064 * BF15
- .575701E-03 * BF16 - .839270E-03 * BF17 + 4.463 * BF18)

Airport: JFK

Basis Functions
===============

BF1 = max(0, CarrierDelay - 0.333);
BF2 = max(0, 0.333 - CarrierDelay);
BF3 = max(0, ScheduleDepartureTime - 49.000);
BF4 = max(0, 49.000 - ScheduleDepartureTime);
BF5 = max(0, GDPHoldingTime - 38.800);
BF6 = max(0, 38.800 - GDPHoldingTime);
BF7 = max(0, ScheduleDepartureTime - 77.000);
BF10 = max(0, 0.500 - SwapAircraftRate) * BF2;
BF11 = max(0, ScheduleDepartureTime - 37.000);
BF13 = max(0, DepartureDemandRatio30 - 7.090) * BF11;
BF14 = max(0, 7.090 - DepartureDemandRatio30) * BF11;
BF15 = max(0, InboundDelay - 74.000) * BF1;
BF16 = max(0, 74.000 - InboundDelay) * BF1;
BF17 = max(0, GDPHoldingTime - .108698E-06) * BF2;
BF18 = max(0, GDPHoldingTime - .108698E-06) * BF1;

sqrt(GeneratedDelay) = max(0, 7.472 + 0.038 * BF1 + 13.612 * BF2 + 0.368 * BF3 - 0.072 * BF4
+ 0.033 * BF5 - 0.052 * BF6 - 0.299 * BF7 - 32.846 * BF10
- 0.132 * BF11 - 0.007 * BF13 - 0.016 * BF14

- .209585E-03 * BF15 + .608024E-03 * BF16 + 0.090 * BF17
- .622994E-03 * BF18)

Airport: LAS

Basis Functions
===============

BF1 = max(0, CarrierDelay - 4.200);
BF2 = max(0, 4.200 - CarrierDelay);
BF3 = max(0, ScheduleDepartureTime - 48.000) * BF2;
BF4 = max(0, 48.000 - ScheduleDepartureTime) * BF2;
BF5 = max(0, GDPHoldingTime - 14.100) * BF2;
BF6 = max(0, 14.100 - GDPHoldingTime) * BF2;
BF7 = max(0, SwapAircraftRate + .980553E-09);
BF8 = max(0, InboundDelay + 10.833) * BF1;
BF9 = max(0, - 10.833 - InboundDelay) * BF1;
BF10 = max(0, CarrierDelay - 2.857) * BF7;
BF11 = max(0, 2.857 - CarrierDelay) * BF7;
BF12 = max(0, ScheduleDepartureTime - 60.000);
BF13 = max(0, 60.000 - ScheduleDepartureTime);
BF14 = max(0, AAR - 12.000);
BF15 = max(0, 12.000 - AAR);
BF16 = max(0, DepartureDemandRatio30 - 1.210);
BF17 = max(0, 1.210 - DepartureDemandRatio30);
BF18 = max(0, GDPHoldingTime - .329794E-08) * BF13;

sqrt(GeneratedDelay) = max(0, 4.197 + 0.093 * BF1 + 0.219 * BF2 - 0.005 * BF3 - 0.012 * BF4
+ 0.010 * BF5 - 0.028 * BF6 + 5.099 * BF7 - .456372E-03 * BF8
+ 0.002 * BF9 - 0.114 * BF10 + 1.745 * BF11
- 0.035 * BF12 - 0.023 * BF13 - 0.120 * BF14
+ 0.024 * BF15 + 0.588 * BF16 + 0.046 * BF17
+ 0.002 * BF18)

Airport: LAX

Basis Functions
===============

BF1 = max(0, CarrierDelay - 9.636);
BF2 = max(0, 9.636 - CarrierDelay);
BF3 = max(0, GDPHoldingTime - 1.200) * BF2;
BF4 = max(0, 1.200 - GDPHoldingTime) * BF2;
BF5 = max(0, SwapAircraftRate - .158304E-10) * BF2;
BF6 = max(0, ScheduleDepartureTime - 39.000) * BF2;
BF7 = max(0, 39.000 - ScheduleDepartureTime) * BF2;
BF8 = max(0, InboundDelay - 55.000);
BF9 = max(0, 55.000 - InboundDelay);
BF10 = max(0, DepartureDemandRatio30 - 0.890);
BF11 = max(0, 0.890 - DepartureDemandRatio30);
BF12 = max(0, InboundDelay - 27.000) * BF1;
BF13 = max(0, 27.000 - InboundDelay) * BF1;
BF14 = max(0, GDPHoldingTime - .104740E-06) * BF9;

BF15 = max(0, SwapAircraftRate - .158304E-10);
BF16 = max(0, InboundDelay - 5.000) * BF10;
BF17 = max(0, 5.000 - InboundDelay) * BF10;
BF18 = max(0, InboundDelay + 27.333) * BF15;

sqrt(GeneratedDelay) = max(0, 3.320 + 0.065 * BF1 - 0.112 * BF2 + 0.006 * BF3 - 0.025 * BF4
+ 0.877 * BF5 - 0.002 * BF6 - 0.006 * BF7 + 0.011 * BF8
+ 0.013 * BF9 + 0.069 * BF10 - 1.515 * BF11
- 0.004 * BF12 + .933777E-03 * BF13 + .718799E-03 * BF14
+ 0.025 * BF16 + 0.038 * BF17 + 0.072 * BF18)

Airport: LGA

Basis Functions
===============

BF1 = max(0, CarrierDelay - 0.333);
BF2 = max(0, 0.333 - CarrierDelay);
BF3 = max(0, ActuralEnrouteWeather - 2459.667);
BF4 = max(0, 2459.667 - ActuralEnrouteWeather);
BF6 = max(0, 4.000 - DepartureDemandRatio30);
BF7 = max(0, GDPHoldingTime - 15.000);
BF8 = max(0, 15.000 - GDPHoldingTime);
BF9 = max(0, ScheduleDepartureTime - 30.000);
BF10 = max(0, 30.000 - ScheduleDepartureTime);
BF11 = max(0, InboundDelay - 155.000) * BF7;
BF12 = max(0, 155.000 - InboundDelay) * BF7;
BF13 = (TerminalWeather = 2 OR TerminalWeather = 4 OR TerminalWeather = 9);
BF16 = max(0, 50.400 - InboundDelay) * BF8;
BF17 = max(0, SwapAircraftRate + .308945E-09) * BF2;
BF18 = max(0, InboundDelay + 35.333) * BF9;

sqrt(GeneratedDelay) = max(0, 9.608 + 0.069 * BF1 - 1.845 * BF2 + .140112E-03 * BF3
- .448500E-03 * BF4 - 1.195 * BF6 - 0.260 * BF10
+ .593182E-03 * BF11 + .336615E-03 * BF12 + 1.095 * BF13
- 0.002 * BF16 + 18.586 * BF17 - .228610E-03 * BF18)

Airport: MCO

Basis Functions
===============

BF1 = max(0, GDPHoldingTime - 12.000);
BF2 = max(0, 12.000 - GDPHoldingTime);
BF3 = max(0, CarrierDelay - 6.875) * BF2;
BF4 = max(0, 6.875 - CarrierDelay) * BF2;
BF6 = max(0, 91.167 - InboundDelay) * BF1;
BF7 = (TerminalWeather = 6 OR TerminalWeather = 8 OR TerminalWeather = 9);
BF8 = (TerminalWeather = 0 OR TerminalWeather = 1 OR TerminalWeather = 2 OR TerminalWeather = 4 OR TerminalWeather = 5 OR TerminalWeather = 7);
BF10 = max(0, 0.500 - SwapAircraftRate) * BF8;
BF12 = max(0, 5.830 - DepartureDemandRatio30);
BF13 = max(0, ScheduleDepartureTime - 72.000);

BF14 = max(0, 72.000 - ScheduleDepartureTime);
BF16 = max(0, 27.000 - CarrierDelay) * BF8;
BF18 = max(0, 61.500 - InboundDelay);

sqrt(GeneratedDelay) = max(0, 9.417 + 0.005 * BF3 - 0.020 * BF4 + .862865E-03 * BF6
- 2.420 * BF7 - 4.881 * BF10 - 0.480 * BF12
- 0.062 * BF13 - 0.021 * BF14 - 0.053 * BF16
+ 0.011 * BF18)

Airport: MDW

Basis Functions
===============

BF1 = max(0, GDPHoldingTime - 15.300);
BF2 = max(0, 15.300 - GDPHoldingTime);
BF3 = max(0, CarrierDelay - 7.500) * BF2;
BF4 = max(0, 7.500 - CarrierDelay) * BF2;
BF5 = max(0, ScheduleDepartureTime - 74.000) * BF2;
BF6 = max(0, 74.000 - ScheduleDepartureTime) * BF2;
BF8 = max(0, 86.000 - InboundDelay) * BF1;
BF10 = max(0, 0.500 - SwapAircraftRate);
BF12 = max(0, 2283.500 - ScheduleEnrouteWeather) * BF10;
BF14 = max(0, 0.714 - SecurityDelay) * BF10;
BF15 = max(0, TurnaroundTime - 30.000) * BF10;
BF16 = max(0, 30.000 - TurnaroundTime) * BF10;
BF17 = max(0, CarrierDelay + .489912E-07) * BF10;
BF18 = max(0, SwapAircraftRate + .122835E-08) * BF1;

sqrt(GeneratedDelay) = max(0, 7.735 - 0.049 * BF2 + 0.002 * BF3 - 0.012 * BF4 - 0.006 * BF5
- 0.002 * BF6 + .795513E-03 * BF8 - .570800E-03 * BF12
- 6.504 * BF14 - 0.007 * BF15 + 0.205 * BF16
+ 0.094 * BF17 - 0.052 * BF18)

Airport: MEM

Basis Functions
===============

BF1 = max(0, GDPHoldingTime - 24.300);
BF2 = max(0, 24.300 - GDPHoldingTime);
BF4 = max(0, 49.000 - CarrierDelay) * BF2;
BF6 = max(0, 0.667 - DepartureDemandRatio_ADR15);
BF7 = max(0, InboundDelay - 73.333) * BF1;
BF9 = (TerminalWeather = 8);
BF11 = max(0, SwapAircraftRate - 0.200) * BF2;
BF14 = max(0, 104.000 - InboundDelay);
BF16 = max(0, 28.400 - GDPHoldingTime) * BF14;
BF17 = max(0, CarrierDelay - 1.000) * BF14;

sqrt(GeneratedDelay) = max(0, 1.456 + 0.050 * BF1 + 0.204 * BF2 - 0.003 * BF4 - 2.115 * BF6
- .294911E-03 * BF7 + 2.966 * BF9 + 0.584 * BF11
+ 0.047 * BF14 - 0.002 * BF16 + .415180E-03 * BF17)

Airport: MIA

Basis Functions
===============

BF1 = max(0, CarrierDelay - 1.500);
BF2 = max(0, 1.500 - CarrierDelay);
BF3 = max(0, GDPHoldingTime - 16.000);
BF4 = max(0, 16.000 - GDPHoldingTime);
BF6 = max(0, 68.000 - ScheduleDepartureTime) * BF2;
BF7 = (TerminalWeather = 8) * BF4;
BF11 = max(0, CarrierDelay - 34.500) * BF4;
BF12 = max(0, 34.500 - CarrierDelay) * BF4;
BF13 = max(0, DepartureDemandRatio30 - 3.000);
BF14 = max(0, 3.000 - DepartureDemandRatio30);
BF15 = max(0, InboundDelay + 25.000);
BF17 = max(0, InboundDelay + 41.000) * BF4;
BF18 = (TerminalWeather = 5 OR TerminalWeather = 8 OR TerminalWeather = 9) * BF4;

sqrt(GeneratedDelay) = max(0, 7.417 + 0.019 * BF1 + 0.054 * BF3 - 0.078 * BF4 - 0.028 * BF6
+ 0.212 * BF7 + 0.002 * BF11 - 0.006 * BF12
+ 0.608 * BF13 - 0.358 * BF14 - 0.056 * BF15
+ 0.003 * BF17 + 0.066 * BF18)

Airport: MSP

Basis Functions
===============

BF1 = max(0, CarrierDelay - 1.000);
BF2 = max(0, 1.000 - CarrierDelay);
BF3 = max(0, GDPHoldingTime - 3.000) * BF2;
BF6 = max(0, 0.950 - DepartureDemandRatio_ADR15);
BF7 = max(0, GDPHoldingTime - 27.700);
BF8 = max(0, 27.700 - GDPHoldingTime);
BF10 = max(0, 21.125 - InboundDelay) * BF1;
BF11 = (TerminalWeather = 0 OR TerminalWeather = 1 OR TerminalWeather = 4 OR TerminalWeather = 5
OR TerminalWeather = 7) * BF8;
BF13 = max(0, ScheduleDepartureTime - 30.000) * BF8;
BF14 = max(0, 30.000 - ScheduleDepartureTime) * BF8;
BF16 = max(0, 73.000 - InboundDelay) * BF6;
BF18 = max(0, 48.600 - CarrierDelay) * BF8;

sqrt(GeneratedDelay) = max(0, 5.913 + 0.023 * BF1 - 0.495 * BF2 + 0.026 * BF3 - 3.418 * BF6
+ 0.032 * BF7 + 0.068 * BF8 + 0.001 * BF10 - 0.047 * BF11
+ .221791E-03 * BF13 - 0.006 * BF14 + 0.028 * BF16
- 0.002 * BF18)

Airport: ORD

Basis Functions
===============

BF1 = max(0, CarrierDelay - 1.000);
BF2 = max(0, 1.000 - CarrierDelay);
BF3 = max(0, GDPHoldingTime - 5.700);
BF4 = max(0, 5.700 - GDPHoldingTime);
BF5 = max(0, ScheduleDepartureTime - 63.000);
BF6 = max(0, 63.000 - ScheduleDepartureTime);
BF7 = max(0, ActuralEnrouteWeather - 62.286);
BF8 = max(0, 62.286 - ActuralEnrouteWeather);
BF9 = max(0, DepartureDemandRatio30 - 0.230) * BF7;
BF10 = max(0, SwapAircraftRate - .696272E-09);
BF11 = (RWY_CODE = 1);
BF13 = max(0, CarrierDelay - 4.333) * BF4;
BF14 = max(0, 4.333 - CarrierDelay) * BF4;
BF15 = (TerminalWeather = 2 OR TerminalWeather = 5 OR TerminalWeather = 7 OR TerminalWeather = 9) * BF7;
BF18 = max(0, 78.200 - InboundDelay) * BF3;

sqrt(GeneratedDelay) = max(0, 4.974 + 0.055 * BF1 - 0.123 * BF2 + 0.022 * BF3 - 0.039 * BF4
- 0.013 * BF5 - 0.034 * BF6 - .119693E-03 * BF7
- 0.004 * BF8 + .368412E-03 * BF9 + 5.283 * BF10
- 0.536 * BF11 + 0.003 * BF13 - 0.033 * BF14
+ .537073E-03 * BF15 + .628346E-03 * BF18)

Airport: PDX

Basis Functions
===============

BF1 = max(0, CarrierDelay - 29.167);
BF2 = max(0, 29.167 - CarrierDelay);
BF3 = max(0, GDPHoldingTime - 17.000) * BF2;
BF4 = max(0, 17.000 - GDPHoldingTime) * BF2;
BF5 = max(0, SwapAircraftRate - .511732E-10) * BF2;
BF6 = max(0, InboundDelay + 1.667);
BF7 = max(0, - 1.667 - InboundDelay);
BF8 = max(0, GDPHoldingTime - 17.000) * BF6;
BF9 = max(0, 17.000 - GDPHoldingTime) * BF6;
BF10 = max(0, InboundDelay - 19.000) * BF2;
BF11 = max(0, 19.000 - InboundDelay) * BF2;
BF13 = max(0, 69.000 - TurnaroundTime);
BF14 = max(0, ScheduleDepartureTime - 58.000) * BF13;
BF15 = max(0, 58.000 - ScheduleDepartureTime) * BF13;
BF16 = max(0, ScheduleDepartureTime - 33.000);
BF17 = max(0, 33.000 - ScheduleDepartureTime);
BF18 = max(0, DepartureDemandRatio30 - 0.270) * BF13;

sqrt(GeneratedDelay) = max(0, 5.902 + 0.054 * BF1 - 0.020 * BF2 + 0.002 * BF3 - 0.007 * BF4
+ 0.306 * BF5 - 0.121 * BF6 + 0.007 * BF7 + .312280E-03 * BF8

+ 0.005 * BF9 + 0.001 * BF10 - .788510E-03 * BF11
+ 0.021 * BF13 - 0.001 * BF14 - .539767E-03 * BF15
+ 0.002 * BF16 + 0.061 * BF17 + 0.006 * BF18)

Airport: PHL

Basis Functions
===============

BF1 = max(0, CarrierDelay - 0.333);
BF2 = max(0, 0.333 - CarrierDelay);
BF3 = max(0, GDPHoldingTime - 19.700);
BF4 = max(0, 19.700 - GDPHoldingTime);
BF6 = max(0, 3.286 - ArrivalDemandRatio30) * BF4;
BF7 = max(0, ActuralEnrouteWeather - 3160.500);
BF8 = max(0, 3160.500 - ActuralEnrouteWeather);
BF9 = max(0, ScheduleDepartureTime - 56.000) * BF4;
BF10 = max(0, 56.000 - ScheduleDepartureTime) * BF4;
BF11 = max(0, DepartureDemandRatio30 - 2.250);
BF12 = max(0, 2.250 - DepartureDemandRatio30);
BF14 = max(0, 34.000 - ScheduleDepartureTime);
BF15 = max(0, ScheduleEnrouteWeather + .147176E-04);
BF17 = max(0, 0.500 - SwapAircraftRate) * BF4;
BF18 = (RWY_CODE = 2) * BF4;

sqrt(GeneratedDelay) = max(0, 9.239 + 0.056 * BF1 - 2.314 * BF2 + 0.024 * BF3 + 0.205 * BF4
- 0.068 * BF6 + .533943E-03 * BF7 - .723781E-03 * BF8
+ 0.004 * BF9 + 0.003 * BF10 + 0.298 * BF11
- 1.044 * BF12 - 0.291 * BF14 - .469390E-03 * BF15
- 0.281 * BF17 - 0.033 * BF18)

Airport: PHX

Basis Functions
===============

BF1 = max(0, CarrierDelay - 2.143);
BF2 = max(0, 2.143 - CarrierDelay);
BF3 = max(0, GDPHoldingTime - 2.300) * BF2;
BF4 = max(0, 2.300 - GDPHoldingTime) * BF2;
BF5 = max(0, SwapAircraftRate - 0.060);
BF6 = max(0, 0.060 - SwapAircraftRate);
BF7 = max(0, AAR - 15.000);
BF8 = max(0, 15.000 - AAR);
BF9 = max(0, ScheduleDepartureTime - 36.000) * BF6;
BF10 = max(0, 36.000 - ScheduleDepartureTime) * BF6;
BF11 = max(0, InboundDelay + 23.500);
BF12 = max(0, - 23.500 - InboundDelay);
BF13 = max(0, DepartureDemandRatio30 - 2.800) * BF11;
BF14 = max(0, 2.800 - DepartureDemandRatio30) * BF11;
BF16 = max(0, 22.375 - CarrierDelay) * BF6;
BF17 = max(0, ScheduleDepartureTime - 51.000);
BF18 = max(0, 51.000 - ScheduleDepartureTime);

sqrt(GeneratedDelay) = max(0, 4.821 + 0.055 * BF1 + 0.038 * BF3 - 0.150 * BF4 + 1.750 * BF5
+ 16.436 * BF6 - 0.204 * BF7 + 0.287 * BF8 - 0.437 * BF9
- 3.114 * BF10 + 0.255 * BF12 + 0.009 * BF13
- 0.007 * BF14 - 1.076 * BF16 + 0.033 * BF17
+ 0.027 * BF18)

Airport: PIT

Basis Functions
===============

BF1 = max(0, GDPHoldingTime - 17.500);
BF2 = max(0, 17.500 - GDPHoldingTime);
BF3 = max(0, CarrierDelay - 0.333) * BF2;
BF4 = max(0, 0.333 - CarrierDelay) * BF2;
BF5 = max(0, InboundDelay - 41.000) * BF1;
BF6 = max(0, 41.000 - InboundDelay) * BF1;
BF8 = max(0, 15.500 - DepartureDemandRatio30) * BF2;
BF9 = max(0, SwapAircraftRate - 0.250);
BF10 = max(0, 0.250 - SwapAircraftRate);
BF12 = (TerminalWeather = 0 OR TerminalWeather = 1 OR TerminalWeather = 2 OR TerminalWeather = 4
OR TerminalWeather = 5);
BF13 = max(0, CarrierDelay - 37.000) * BF12;
BF14 = max(0, 37.000 - CarrierDelay) * BF12;
BF15 = max(0, InboundDelay + 6.000) * BF10;
BF17 = max(0, InboundDelay + 4.500) * BF2;
BF18 = max(0, - 4.500 - InboundDelay) * BF2;

sqrt(GeneratedDelay) = max(0, 9.473 + 0.038 * BF1 + 0.163 * BF2 + 0.003 * BF3 - 0.149 * BF4
- .195893E-03 * BF5 + .653026E-03 * BF6 - 0.021 * BF8
+ 4.782 * BF9 - 8.473 * BF10 - 0.035 * BF13
- 0.062 * BF14 - 0.102 * BF15 + 0.001 * BF17
- 0.002 * BF18)

Airport: SAN

Basis Functions
===============
BF2 = max(0, 1.333 - CarrierDelay);
BF6 = max(0, 1200.000 - ICM_CR);
BF7 = max(0, VISIB_CR - 6.000) * BF6;
BF9 = max(0, SwapAircraftRate - .297868E-10) * BF2;
BF10 = max(0, ScheduleDepartureTime - 43.000);
BF13 = max(0, 4.500 - DepartureDemandRatio15) * BF2;
BF14 = max(0, GDPHoldingTime - 8.500);
BF15 = max(0, 8.500 - GDPHoldingTime);
BF17 = max(0, 55.333 - InboundDelay) * BF14;
BF18 = max(0, CarrierDelay - .113268E-06) * BF15;

sqrt(GeneratedDelay) = max(0, 4.035 + 0.003 * BF6 - .864972E-03 * BF7 + 7.185 * BF9
- 0.013 * BF10 - 0.144 * BF13 - 0.166 * BF15
+ 0.001 * BF17 + 0.011 * BF18)

Airport: SEA

Basis Functions
===============

```
BF1 = max(0, CarrierDelay - 10.200);
BF2 = max(0, 10.200 - CarrierDelay );
BF4 = max(0, 53.000 - GDPHoldingTime ) * BF2;
BF5 = max(0, InboundDelay - 113.000);
BF6 = max(0, 113.000 - InboundDelay );
BF7 = max(0, DepartureDemandRatio30 - 2.470);
BF8 = max(0, 2.470 - DepartureDemandRatio30 );
BF9 = max(0, SwapAircraftRate - 0.250) * BF2;
BF10 = max(0, 0.250 - SwapAircraftRate ) * BF2;
BF12 = max(0, 999.000 - Distance ) * BF6;
BF14 = max(0, 72.000 - CarrierDelay ) * BF6;
BF15 = max(0, ScheduleDepartureTime - 43.000);
BF16 = max(0, 43.000 - ScheduleDepartureTime );
BF17 = max(0, GDPHoldingTime - .179148E-07) * BF6;
BF18 = max(0, GDPHoldingTime - .179148E-07);

sqrt(GeneratedDelay) = max(0, 3.021 + 0.040 * BF1 + 0.537 * BF2 - 0.007 * BF4 + 0.021 * BF5
         + 0.040 * BF6 + 0.774 * BF7 - 0.304 * BF8 + 0.391 * BF9
         - 1.111 * BF10 - .100473E-04 * BF12 - .395528E-03 * BF14
         - 0.011 * BF15 - 0.027 * BF16 + .674974E-03 * BF17
         - 0.053 * BF18)
```

Airport: SFO

Basis Functions
===============

```
BF1 = max(0, CarrierDelay - 1.750);
BF2 = max(0, 1.750 - CarrierDelay );
BF3 = max(0, GDPHoldingTime - 0.700);
BF4 = max(0, 0.700 - GDPHoldingTime );
BF5 = max(0, DepartureDemandRatio_ADR15 - 0.875) * BF2;
BF6 = max(0, 0.875 - DepartureDemandRatio_ADR15 ) * BF2;
BF7 = max(0, InboundDelay - 100.500) * BF1;
BF8 = max(0, 100.500 - InboundDelay ) * BF1;
BF10 = max(0, 0.500 - SwapAircraftRate );
BF11 = max(0, LegNumber - 2.000) * BF10;
BF12 = max(0, 2.000 - LegNumber ) * BF10;
BF13 = max(0, CarrierDelay - 59.000) * BF4;
BF14 = max(0, 59.000 - CarrierDelay ) * BF4;
BF15 = max(0, Distance - 2470.000);
BF16 = max(0, 2470.000 - Distance );
BF17 = max(0, InboundDelay - 36.857) * BF3;
BF18 = max(0, 36.857 - InboundDelay ) * BF3;

sqrt(GeneratedDelay) = max(0, 7.374 - 0.072 * BF1 - 0.145 * BF2 + 0.027 * BF3 + 4.739 * BF4
         - 0.164 * BF5 - 0.850 * BF6 + .954112E-03 * BF7
         + 0.001 * BF8 - 7.102 * BF10 - 0.323 * BF11
```

- 1.188 * BF12 - 0.044 * BF13 - 0.089 * BF14
- 0.006 * BF15 - .293950E-03 * BF16 - .149685E-03 * BF17
+ 0.001 * BF18)

Airport: SLC

Basis Functions
===============

BF1 = max(0, CarrierDelay - 1.000);
BF2 = max(0, 1.000 - CarrierDelay);
BF3 = max(0, GDPHoldingTime - 25.000) * BF2;
BF4 = max(0, 25.000 - GDPHoldingTime) * BF2;
BF5 = max(0, DepartureDemandRatio_ADR15 - 1.458) * BF2;
BF6 = max(0, 1.458 - DepartureDemandRatio_ADR15) * BF2;
BF7 = max(0, ArrivalDemandRatio_AAR30 + .947311E-08) * BF2;
BF9 = max(0, 0.200 - SwapAircraftRate) * BF2;
BF11 = max(0, 1.000 - InboundDelay) * BF1;
BF12 = max(0, InboundDelay + 3.333);
BF14 = max(0, GDPHoldingTime - 2.200) * BF12;
BF15 = max(0, 2.200 - GDPHoldingTime) * BF12;
BF17 = max(0, 23.400 - GDPHoldingTime);
BF18 = max(0, InboundDelay + 31.000) * BF2;

sqrt(GeneratedDelay) = max(0, 5.293 + 0.063 * BF1 + 3.676 * BF2 + 0.069 * BF3 - 0.073 * BF4
- 1.335 * BF5 - 0.976 * BF6 + 1.101 * BF7 - 16.245 * BF9
+ 0.003 * BF11 - 0.057 * BF12 - .416106E-03 * BF14
+ 0.016 * BF15 - 0.084 * BF17 + 0.018 * BF18)

Airport: STL

Basis Functions
===============

BF1 = max(0, GDPHoldingTime - 10.000);
BF2 = max(0, 10.000 - GDPHoldingTime);
BF3 = max(0, CarrierDelay - 7.500) * BF2;
BF4 = max(0, 7.500 - CarrierDelay) * BF2;
BF6 = max(0, 103.333 - InboundDelay) * BF1;
BF7 = max(0, DepartureDemandRatio30 - 1.670) * BF2;
BF8 = max(0, 1.670 - DepartureDemandRatio30) * BF2;
BF9 = max(0, ScheduleDepartureTime - 65.000);
BF10 = max(0, 65.000 - ScheduleDepartureTime);
BF12 = max(0, 58.500 - CarrierDelay);
BF15 = max(0, InboundDelay + 4.333);
BF17 = max(0, InboundDelay - 1.667) * BF2;
BF18 = max(0, 1.667 - InboundDelay) * BF2;

sqrt(GeneratedDelay) = max(0, 6.445 + 0.024 * BF1 + 0.005 * BF3 - 0.023 * BF4 + .449881E-03 * BF6
+ 0.081 * BF7 - 0.047 * BF8 - 0.023 * BF9 - 0.018 * BF10
- 0.035 * BF12 - 0.028 * BF15 + 0.002 * BF17
- 0.004 * BF18)

Airport: TPA

Basis Functions
===============

BF1 = max(0, GDPHoldingTime - 7.000);
BF2 = max(0, 7.000 - GDPHoldingTime);
BF3 = max(0, CarrierDelay - 5.000) * BF2;
BF4 = max(0, 5.000 - CarrierDelay) * BF2;
BF5 = max(0, InboundDelay - 96.000) * BF1;
BF6 = max(0, 96.000 - InboundDelay) * BF1;
BF7 = max(0, DepartureDemandRatio30 - 7.000);
BF8 = max(0, 7.000 - DepartureDemandRatio30);
BF9 = max(0, ScheduleDepartureTime - 71.000);
BF10 = max(0, 71.000 - ScheduleDepartureTime);
BF11 = max(0, SwapAircraftRate - 0.500) * BF2;
BF12 = max(0, 0.500 - SwapAircraftRate) * BF2;
BF13 = max(0, DepartureDemandRatio30 - 1.450) * BF2;
BF14 = max(0, 1.450 - DepartureDemandRatio30) * BF2;
BF16 = max(0, 61.000 - CarrierDelay) * BF8;
BF17 = max(0, InboundDelay + .133610E-06);
BF18 = max(0, - .133610E-06 - InboundDelay);

sqrt(GeneratedDelay) = max(0, 8.117 + 0.188 * BF2 + 0.006 * BF3 - 0.033 * BF4 + .261904E-03 * BF5
+ .684080E-03 * BF6 - 1.051 * BF7 - 0.180 * BF8
- 0.092 * BF9 - 0.027 * BF10 + 1.161 * BF11
- 0.541 * BF12 + 0.061 * BF13 + 0.071 * BF14
- 0.008 * BF16 - 0.008 * BF17 - 0.022 * BF18)

B.2 Airport Absorbed Delay Models

Airport: ATL

Basis Functions
===============

BF1 = max(0, InboundDelay - 18.222);
BF2 = max(0, 18.222 - InboundDelay);
BF3 = max(0, TurnaroundTime - 84.500);
BF4 = max(0, 84.500 - TurnaroundTime);
BF5 = max(0, InboundDelay - 21.000) * BF4;
BF6 = max(0, 21.000 - InboundDelay) * BF4;
BF7 = max(0, CarrierDelay - 39.667);
BF8 = max(0, 39.667 - CarrierDelay);
BF10 = max(0, 54.700 - GDPHoldingTime);
BF11 = max(0, NumberSeats - 161.400);
BF12 = max(0, 161.400 - NumberSeats);
BF13 = max(0, InboundDelay - 69.333) * BF10;

```
 BF14 = max(0, 69.333 - InboundDelay ) * BF10;
 BF15 = max(0, InboundDelay - 8.750) * BF8;
 BF16 = max(0, 8.750 - InboundDelay ) * BF8;
 BF17 = max(0, InboundDelay + .441525E-06) * BF12;
 BF18 = max(0, - .441525E-06 - InboundDelay ) * BF12;

sqrt(AbsorbedDelay) = min(0, -1.324 - 0.027 * BF1 + 0.053 * BF2 + 0.006 * BF3 + 0.038 * BF4
           + .480074E-03 * BF5 - 0.001 * BF6 + 0.004 * BF7
           - 0.033 * BF8 - 0.025 * BF10 + 0.013 * BF11
           + 0.002 * BF12 + .596830E-03 * BF13 + .154085E-03 * BF14
           + .208147E-03 * BF15 + 0.001 * BF16 - .253714E-03 * BF17
           - .252755E-03 * BF18)
```

Airport: BOS

```
 Basis Functions
 ===============

 BF1 = max(0, InboundDelay - 7.000);
 BF2 = max(0, 7.000 - InboundDelay );
 BF4 = max(0, 82.500 - GDPHoldingTime );
 BF6 = max(0, 122.800 - TurnaroundTime ) * BF1;
 BF7 = max(0, Distance - 300.000) * BF1;
 BF10 = max(0, 1.000 - CarrierDelay );
 BF11 = max(0, InboundDelay - 57.000) * BF4;
 BF14 = max(0, 74.750 - TurnaroundTime ) * BF4;
 BF15 = max(0, InboundDelay + 10.500) * BF10;
 BF16 = max(0, - 10.500 - InboundDelay ) * BF10;

sqrt(AbsorbedDelay) = min(0, -0.270 - 0.051 * BF1 + 0.077 * BF2 - 0.025 * BF4 + .470724E-03 * BF6
           + .120250E-04 * BF7 + .213996E-03 * BF11 + .111745E-03 * BF14
           - 0.012 * BF15 - 0.073 * BF16)
```

Airport: BWI

```
 Basis Functions
 ===============

 BF1 = max(0, InboundDelay - 7.500);
 BF2 = max(0, 7.500 - InboundDelay );
 BF3 = max(0, TurnaroundTime - 95.750) * BF1;
 BF5 = max(0, Distance - 1043.250);
 BF6 = max(0, 1043.250 - Distance );
 BF7 = max(0, CarrierDelay - 1.000);
 BF8 = max(0, 1.000 - CarrierDelay );
 BF9 = max(0, GDPHoldingTime - 36.000) * BF6;
 BF10 = max(0, 36.000 - GDPHoldingTime ) * BF6;
 BF12 = max(0, 71.660 - TurnaroundTime );
 BF13 = max(0, InboundDelay + 11.000) * BF12;
 BF14 = max(0, - 11.000 - InboundDelay ) * BF12;
 BF15 = max(0, NumberSeats - 108.000) * BF1;
 BF18 = max(0, 4.000 - InboundDelay ) * BF8;
```

sqrt(AbsorbedDelay) = min(0, -2.423 - 0.045 * BF1 + 0.081 * BF2 + .716995E-03 * BF3
+ .880120E-03 * BF5 + 0.002 * BF6 + 0.005 * BF7
- 0.501 * BF8 + .103871E-04 * BF9 - .360186E-04 * BF10
+ 0.007 * BF12 + .953999E-03 * BF13 - 0.004 * BF14
+ .314077E-03 * BF15 + 0.039 * BF18)

Airport: CLE

Basis Functions
===============

BF1 = max(0, InboundDelay - 6.000);
BF2 = max(0, 6.000 - InboundDelay);
BF4 = max(0, 71.660 - TurnaroundTime);
BF5 = max(0, InboundDelay + 11.000) * BF4;
BF6 = max(0, - 11.000 - InboundDelay) * BF4;
BF7 = max(0, ScheduleDepartureTime - 72.000) * BF4;
BF9 = max(0, ArrivalDemandRatio_AAR30_Dest - 0.333);
BF10 = max(0, 0.333 - ArrivalDemandRatio_AAR30_Dest);
BF11 = max(0, GDPHoldingTime + .661021E-08);
BF13 = max(0, 0.500 - CarrierDelay);
BF14 = max(0, Weight - 4.090);
BF15 = max(0, 4.090 - Weight);
BF17 = max(0, 15.000 - InboundDelay) * BF13;
BF18 = max(0, ScheduleDepartureTime - 24.000) * BF9;

sqrt(AbsorbedDelay) = min(0, -2.676 - 0.039 * BF1 + 0.078 * BF2 + 0.017 * BF4 + .908381E-03 * BF5
- 0.004 * BF6 - 0.002 * BF7 - 1.004 * BF9 + 5.281 * BF10
+ 0.012 * BF11 - 1.655 * BF13 + 0.638 * BF14
+ 3.632 * BF15 + 0.061 * BF17 + 0.025 * BF18)

Airport: CLT

Basis Functions
===============

BF1 = max(0, InboundDelay - 18.444);
BF2 = max(0, 18.444 - InboundDelay);
BF3 = max(0, GDPHoldingTime - 17.000);
BF4 = max(0, 17.000 - GDPHoldingTime);
BF6 = max(0, 91.140 - TurnaroundTime) * BF1;
BF8 = max(0, 6.000 - CarrierDelay) * BF4;
BF10 = max(0, 158.120 - NumberSeats) * BF1;
BF12 = max(0, 58.710 - TurnaroundTime);
BF14 = max(0, 60.000 - TurnaroundTime) * BF2;
BF16 = max(0, 5.667 - CarrierDelay) * BF2;
BF18 = max(0, 4.667 - CarrierDelay);

sqrt(AbsorbedDelay) = min(0, -2.118 - 0.043 * BF1 + 0.049 * BF2 + 0.009 * BF3 + 0.001 * BF6
- 0.006 * BF8 - .297411E-03 * BF10 + 0.036 * BF12
- 0.001 * BF14 + 0.008 * BF16 - 0.163 * BF18)

Airport: CVG

Basis Functions
===============

BF1 = max(0, InboundDelay - 20.000);
BF2 = max(0, 20.000 - InboundDelay);
BF3 = max(0, TurnaroundTime - 104.330) * BF1;
BF4 = max(0, 104.330 - TurnaroundTime) * BF1;
BF5 = max(0, TurnaroundTime - 70.700);
BF6 = max(0, 70.700 - TurnaroundTime);
BF7 = max(0, GDPHoldingTime - .816149E-07);
BF8 = max(0, CarrierDelay - 0.500);
BF9 = max(0, 0.500 - CarrierDelay);
BF10 = max(0, InboundDelay - 21.400) * BF6;
BF11 = max(0, 21.400 - InboundDelay) * BF6;
BF12 = max(0, InboundDelay + 12.000) * BF9;
BF13 = max(0, - 12.000 - InboundDelay) * BF9;
BF14 = max(0, NumberSeats - 178.000) * BF1;
BF15 = max(0, 178.000 - NumberSeats) * BF1;
BF16 = max(0, InboundDelay - 4.600) * BF9;
BF18 = max(0, GDPHoldingTime - .816149E-07) * BF2;

sqrt(AbsorbedDelay) = min(0, -3.300 - 0.041 * BF1 + 0.078 * BF2 + .497451E-03 * BF3
+ 0.002 * BF4 + 0.001 * BF5 + 0.037 * BF6 + 0.017 * BF7
+ 0.012 * BF8 + 0.623 * BF9 - 0.002 * BF10
- .967181E-03 * BF11 - 0.082 * BF12 - 0.138 * BF13
- 0.002 * BF14 - .361987E-03 * BF15 + 0.065 * BF16
- .343876E-03 * BF18)

Airport: DCA

Basis Functions
===============

BF1 = max(0, InboundDelay - 8.000);
BF2 = max(0, 8.000 - InboundDelay);
BF3 = max(0, TurnaroundTime - 70.000) * BF1;
BF4 = max(0, 70.000 - TurnaroundTime) * BF1;
BF5 = max(0, CarrierDelay - 4.500);
BF6 = max(0, 4.500 - CarrierDelay);
BF7 = max(0, TurnaroundTime - 61.250);
BF8 = max(0, 61.250 - TurnaroundTime);
BF9 = max(0, NumberSeats - 128.500) * BF8;
BF10 = max(0, 128.500 - NumberSeats) * BF8;
BF11 = max(0, TurnaroundTime - 64.000) * BF2;
BF12 = max(0, 64.000 - TurnaroundTime) * BF2;
BF14 = max(0, 85.000 - GDPHoldingTime);
BF15 = max(0, InboundDelay + 9.333) * BF6;
BF16 = max(0, - 9.333 - InboundDelay) * BF6;
BF17 = max(0, InboundDelay - 32.000) * BF14;

sqrt(AbsorbedDelay) = min(0, -1.241 - 0.022 * BF1 + 0.130 * BF2 - .259964E-03 * BF3

+ .473242E-03 * BF4 + 0.008 * BF5 - 0.052 * BF6
+ 0.004 * BF7 + 0.049 * BF8 + 0.002 * BF9 - .183185E-03 * BF10
- .166914E-03 * BF11 - 0.003 * BF12 - 0.016 * BF14
- 0.003 * BF15 - 0.014 * BF16 + .266873E-03 * BF17)

Airport: DEN

Basis Functions
===============

BF1 = max(0, InboundDelay - 8.000);
BF2 = max(0, 8.000 - InboundDelay);
BF3 = max(0, TurnaroundTime - 102.250) * BF1;
BF4 = max(0, 102.250 - TurnaroundTime) * BF1;
BF6 = max(0, 62.250 - TurnaroundTime);
BF7 = max(0, CarrierDelay - 5.200);
BF8 = max(0, 5.200 - CarrierDelay);
BF9 = max(0, NumberSeats - 108.000) * BF1;
BF10 = max(0, 108.000 - NumberSeats) * BF1;
BF11 = max(0, InboundDelay + 11.700) * BF6;
BF12 = max(0, - 11.700 - InboundDelay) * BF6;
BF13 = max(0, NumberSeats - 110.660);
BF14 = max(0, 110.660 - NumberSeats);
BF15 = max(0, GDPHoldingTime - .283873E-07) * BF6;
BF17 = max(0, 22.000 - InboundDelay) * BF8;

sqrt(AbsorbedDelay) = min(0, -2.135 - 0.071 * BF1 + 0.061 * BF2 + .414740E-03 * BF3
+ .827640E-03 * BF4 + 0.013 * BF6 + 0.008 * BF7
- 0.124 * BF8 + .256014E-03 * BF9 + 0.002 * BF10
+ .800474E-03 * BF11 - 0.004 * BF12 + 0.002 * BF13
+ 0.015 * BF14 + .843695E-03 * BF15 + 0.003 * BF17)

Airport: DFW

Basis Functions
===============

BF1 = max(0, InboundDelay - 3.714);
BF2 = max(0, 3.714 - InboundDelay);
BF3 = max(0, TurnaroundTime - 73.500);
BF4 = max(0, 73.500 - TurnaroundTime);
BF5 = max(0, InboundDelay + 10.000) * BF4;
BF6 = max(0, - 10.000 - InboundDelay) * BF4;
BF7 = max(0, GDPHoldingTime + .101890E-06);
BF8 = max(0, NumberSeats - 92.000);
BF9 = max(0, CarrierDelay - 0.750) * BF8;
BF10 = max(0, 0.750 - CarrierDelay) * BF8;
BF12 = max(0, 2.600 - CarrierDelay) * BF1;
BF13 = max(0, NumberSeats - 170.500) * BF1;
BF15 = max(0, InboundDelay - 99.000) * BF3;
BF16 = max(0, 99.000 - InboundDelay) * BF3;
BF17 = max(0, InboundDelay - 64.500) * BF8;
BF18 = max(0, 64.500 - InboundDelay) * BF8;

sqrt(AbsorbedDelay) = min(0, -2.813 - 0.048 * BF1 + 0.115 * BF2 + 0.014 * BF3 + 0.015 * BF4
+ .824261E-03 * BF5 - 0.003 * BF6 + 0.012 * BF7
+ 0.027 * BF8 + .204332E-03 * BF9 - 0.004 * BF10
+ 0.004 * BF12 - .774182E-03 * BF13 - .319299E-03 * BF15
- .114874E-03 * BF16 + .525390E-03 * BF17 - .288912E-03 * BF18)

Airport: DTW

Basis Functions
===============

BF1 = max(0, InboundDelay - 16.000);
BF4 = max(0, 103.330 - TurnaroundTime) * BF1;
BF5 = max(0, GDPHoldingTime - 50.000);
BF6 = max(0, 50.000 - GDPHoldingTime);
BF7 = max(0, CarrierDelay - 8.833);
BF8 = max(0, 8.833 - CarrierDelay);
BF9 = max(0, TurnaroundTime - 85.660);
BF10 = max(0, 85.660 - TurnaroundTime);
BF11 = max(0, InboundDelay - 17.500) * BF10;
BF12 = max(0, 17.500 - InboundDelay) * BF10;
BF14 = max(0, 31.600 - InboundDelay) * BF8;
BF15 = max(0, InboundDelay + 12.000);
BF17 = max(0, InboundDelay - 12.250) * BF6;
BF18 = max(0, 12.250 - InboundDelay) * BF6;

sqrt(AbsorbedDelay) = min(0, 0.336 + 0.032 * BF1 + 0.001 * BF4 + 0.005 * BF5 - 0.027 * BF6
+ 0.006 * BF7 - 0.134 * BF8 + 0.003 * BF9 + 0.037 * BF10
- 0.001 * BF11 - 0.001 * BF12 + 0.003 * BF14
- 0.082 * BF15 + .189393E-03 * BF17 + .473089E-03 * BF18)

Airport: EWR

Basis Functions
===============

BF1 = max(0, InboundDelay - 10.000);
BF2 = max(0, 10.000 - InboundDelay);
BF3 = max(0, TurnaroundTime - 87.000) * BF1;
BF4 = max(0, 87.000 - TurnaroundTime) * BF1;
BF5 = max(0, CarrierDelay - 0.800);
BF6 = max(0, 0.800 - CarrierDelay);
BF7 = max(0, DepartureDemandRatio30 - 1.340) * BF1;
BF8 = max(0, 1.340 - DepartureDemandRatio30) * BF1;
BF9 = max(0, GDPHoldingTime - .778340E-08);
BF10 = max(0, TurnaroundTime - 75.000);
BF11 = max(0, 75.000 - TurnaroundTime);
BF12 = max(0, NumberSeats - 125.000);
BF14 = max(0, InboundDelay - 65.000) * BF6;
BF15 = max(0, 65.000 - InboundDelay) * BF6;
BF16 = max(0, TurnaroundTime - 76.250) * BF2;
BF17 = max(0, 76.250 - TurnaroundTime) * BF2;
BF18 = max(0, InboundDelay + 27.000) * BF12;

sqrt(AbsorbedDelay) = min(0, -2.756 - 0.021 * BF1 + 0.106 * BF2 - .108494E-03 * BF3
+ .513454E-03 * BF4 + 0.011 * BF5 - 1.445 * BF6
+ 0.002 * BF7 - 0.018 * BF8 + 0.012 * BF9 + 0.006 * BF10
+ 0.034 * BF11 + 0.011 * BF14 + 0.017 * BF15
- .266450E-03 * BF16 - 0.002 * BF17 + .210935E-03 * BF18)

Airport: FLL

Basis Functions
===============

BF1 = max(0, InboundDelay - 5.500);
BF2 = max(0, 5.500 - InboundDelay);
BF4 = max(0, 73.000 - TurnaroundTime) * BF1;
BF5 = max(0, TurnaroundTime - 67.500);
BF6 = max(0, 67.500 - TurnaroundTime);
BF8 = max(0, 5.000 - CarrierDelay);
BF9 = max(0, ScheduleDepartureTime - 45.000) * BF6;
BF10 = max(0, 45.000 - ScheduleDepartureTime) * BF6;
BF12 = max(0, 22.000 - InboundDelay) * BF8;
BF13 = max(0, GDPHoldingTime - 35.000) * BF8;
BF14 = max(0, 35.000 - GDPHoldingTime) * BF8;
BF15 = max(0, NumberSeats - 143.250) * BF1;
BF16 = max(0, 143.250 - NumberSeats) * BF1;
BF18 = max(0, - 7.000 - InboundDelay) * BF6;

sqrt(AbsorbedDelay) = min(0, -1.515 - 0.042 * BF1 + 0.043 * BF2 + 0.001 * BF4 + 0.001 * BF5
+ 0.030 * BF6 - 0.151 * BF8 - .305965E-03 * BF9
- 0.003 * BF10 + 0.008 * BF12 + 0.002 * BF13
- 0.003 * BF14 + .273304E-03 * BF15 - .710284E-03 * BF16
- 0.002 * BF18)

Airport: IAD

Basis Functions
===============

BF1 = max(0, InboundDelay - 14.000);
BF2 = max(0, 14.000 - InboundDelay);
BF3 = max(0, TurnaroundTime - 70.750) * BF1;
BF4 = max(0, 70.750 - TurnaroundTime) * BF1;
BF5 = max(0, CarrierDelay - 1.000);
BF6 = max(0, 1.000 - CarrierDelay);
BF8 = max(0, 86.000 - GDPHoldingTime);
BF9 = max(0, LegNumber - 5.220) * BF1;
BF10 = max(0, 5.220 - LegNumber) * BF1;
BF11 = max(0, TurnaroundTime - 52.500) * BF8;
BF12 = max(0, 52.500 - TurnaroundTime) * BF8;
BF13 = max(0, TurnaroundTime - 56.000) * BF2;
BF14 = max(0, 56.000 - TurnaroundTime) * BF2;
BF15 = max(0, NumberSeats - 110.000) * BF8;
BF16 = max(0, 110.000 - NumberSeats) * BF8;
BF17 = max(0, DepartureDemandRatio30 - 0.710) * BF1;

BF18 = max(0, 0.710 - DepartureDemandRatio30) * BF1;

sqrt(AbsorbedDelay) = min(0, -0.798 - 0.010 * BF1 + 0.077 * BF2 + .344490E-04 * BF3
+ .645234E-03 * BF4 + 0.009 * BF5 - 0.286 * BF6
- 0.022 * BF8 - 0.014 * BF9 - 0.004 * BF10
- .124021E-03 * BF11 + .652704E-03 * BF12 + .668317E-03 * BF13
- 0.002 * BF14 + .731895E-04 * BF15 - .199855E-04 * BF16
+ 0.001 * BF17 - 0.019 * BF18)

Airport: IAH

Basis Functions
===============

BF1 = max(0, InboundDelay - 14.500);
BF2 = max(0, 14.500 - InboundDelay);
BF6 = max(0, 64.300 - GDPHoldingTime);
BF8 = max(0, 74.660 - TurnaroundTime);
BF9 = max(0, InboundDelay + 10.000) * BF8;
BF10 = max(0, - 10.000 - InboundDelay) * BF8;
BF11 = max(0, NumberSeats - 90.000);
BF12 = max(0, 90.000 - NumberSeats);
BF14 = max(0, 7.125 - CarrierDelay) * BF11;
BF15 = max(0, InboundDelay - .257915E-07) * BF12;
BF16 = max(0, .257915E-07 - InboundDelay) * BF12;
BF18 = max(0, 14.000 - InboundDelay) * BF6;

sqrt(AbsorbedDelay) = min(0, -1.779 - 0.029 * BF1 + 0.047 * BF2 - 0.024 * BF6 + 0.001 * BF9
- 0.002 * BF10 + 0.012 * BF11 + 0.036 * BF12
- 0.001 * BF14 - 0.001 * BF15 - 0.002 * BF16
+ .791638E-03 * BF18)

Airport: JFK

Basis Functions
===============

BF1 = max(0, InboundDelay - 12.500);
BF2 = max(0, 12.500 - InboundDelay);
BF3 = max(0, TurnaroundTime - 92.500) * BF1;
BF4 = max(0, 92.500 - TurnaroundTime) * BF1;
BF5 = max(0, CarrierDelay - 9.000);
BF6 = max(0, 9.000 - CarrierDelay);
BF7 = max(0, InboundDelay - 81.000) * BF6;
BF8 = max(0, 81.000 - InboundDelay) * BF6;
BF9 = max(0, TurnaroundTime - 135.000);
BF10 = max(0, 135.000 - TurnaroundTime);
BF11 = max(0, InboundDelay - 90.000) * BF10;
BF12 = max(0, 90.000 - InboundDelay) * BF10;
BF13 = max(0, InboundDelay + 12.000);
BF15 = max(0, InboundDelay - 78.000) * BF9;
BF16 = max(0, 78.000 - InboundDelay) * BF9;
BF17 = max(0, InboundDelay - 87.000);

sqrt(AbsorbedDelay) = min(0, -0.023 + 0.045 * BF1 + 0.028 * BF2 + .164685E-03 * BF3
+ .206367E-03 * BF4 + 0.006 * BF5 - 0.285 * BF6
+ 0.004 * BF7 + 0.003 * BF8 + 0.016 * BF9 + 0.077 * BF10
- .264400E-03 * BF11 - .838000E-03 * BF12 - 0.110 * BF13
- 0.002 * BF15 - .191659E-03 * BF16 + 0.058 * BF17)

Airport: LAS

Basis Functions
===============

BF1 = max(0, InboundDelay - 9.333);
BF2 = max(0, 9.333 - InboundDelay);
BF3 = max(0, TurnaroundTime - 87.160);
BF4 = max(0, 87.160 - TurnaroundTime);
BF5 = max(0, ScheduleDepartureTime - 59.000) * BF4;
BF6 = max(0, 59.000 - ScheduleDepartureTime) * BF4;
BF7 = max(0, InboundDelay - 7.333) * BF4;
BF8 = max(0, 7.333 - InboundDelay) * BF4;
BF9 = max(0, CarrierDelay - 0.333);
BF10 = max(0, 0.333 - CarrierDelay);
BF11 = max(0, NumberSeats - 115.000) * BF1;
BF12 = max(0, 115.000 - NumberSeats) * BF1;
BF13 = max(0, InboundDelay + 8.500) * BF10;
BF14 = max(0, - 8.500 - InboundDelay) * BF10;
BF15 = max(0, NumberSeats - 166.000) * BF2;
BF16 = max(0, 166.000 - NumberSeats) * BF2;
BF17 = max(0, CarrierDelay - 0.600) * BF2;
BF18 = max(0, 0.600 - CarrierDelay) * BF2;

sqrt(AbsorbedDelay) = min(0, -3.487 - 0.048 * BF1 + 0.123 * BF2 + 0.026 * BF3 + 0.047 * BF4
- .543967E-03 * BF5 - .603585E-03 * BF6 + .761984E-03 * BF7
- 0.002 * BF8 + 0.015 * BF9 - 0.930 * BF10
+ .374798E-03 * BF11 - 0.007 * BF12 - 0.019 * BF13
- 0.210 * BF14 + 0.001 * BF15 + .246537E-03 * BF16
- .713822E-03 * BF17 + 0.033 * BF18)

Airport: LAX

Basis Functions
===============

BF1 = max(0, InboundDelay - 12.667);
BF2 = max(0, 12.667 - InboundDelay);
BF4 = max(0, 85.250 - TurnaroundTime) * BF1;
BF5 = max(0, TurnaroundTime - 83.250);
BF6 = max(0, 83.250 - TurnaroundTime);
BF7 = max(0, CarrierDelay - 0.200);
BF8 = max(0, 0.200 - CarrierDelay);
BF9 = max(0, InboundDelay + 9.000) * BF8;
BF10 = max(0, - 9.000 - InboundDelay) * BF8;
BF11 = max(0, DepartureRatio_15 - 0.500) * BF6;
BF12 = max(0, 0.500 - DepartureRatio_15) * BF6;

```
 BF14 = max(0, 0.693 - DepartureDemandRatio_ADR15 );
 BF15 = max(0, NumberSeats - 52.500);
 BF17 = max(0, Distance - 277.800);
 BF18 = max(0, 277.800 - Distance );

sqrt(AbsorbedDelay) = min(0, -2.839 - 0.047 * BF1 + 0.066 * BF2 + 0.001 * BF4 + 0.001 * BF5
          + 0.014 * BF6 + 0.010 * BF7 - 0.090 * BF9 - 0.280 * BF10
          - 0.018 * BF11 + 0.040 * BF12 - 1.223 * BF14
          + 0.006 * BF15 - .217227E-03 * BF17 + 0.005 * BF18)

Airport: LGA

 Basis Functions
 ===============

 BF1 = max(0, InboundDelay - 12.500);
 BF2 = max(0, 12.500 - InboundDelay );
 BF3 = max(0, TurnaroundTime - 77.000) * BF1;
 BF5 = max(0, CarrierDelay - 3.000);
 BF6 = max(0, 3.000 - CarrierDelay );
 BF8 = max(0, 74.000 - TurnaroundTime );
 BF9 = max(0, Distance - 616.370) * BF8;
 BF10 = max(0, 616.370 - Distance ) * BF8;
 BF11 = max(0, GDPHoldingTime - 83.000) * BF6;
 BF12 = max(0, 83.000 - GDPHoldingTime ) * BF6;
 BF13 = max(0, NumberSeats - 73.000);
 BF14 = max(0, 73.000 - NumberSeats );
 BF15 = max(0, InboundDelay + 13.000) * BF8;
 BF16 = max(0, - 13.000 - InboundDelay ) * BF8;
 BF17 = max(0, NumberSeats - 130.000) * BF1;

sqrt(AbsorbedDelay) = min(0, -3.310 - 0.023 * BF1 + 0.112 * BF2 - .107584E-03 * BF3
          + 0.016 * BF5 + 0.225 * BF6 + .586672E-04 * BF9
          + .336713E-04 * BF10 + 0.009 * BF11 - 0.004 * BF12
          + 0.004 * BF13 + 0.043 * BF14 + .660217E-03 * BF15
          - 0.007 * BF16 + .219447E-03 * BF17)

Airport: MCO

 Basis Functions
 ===============

 BF1 = max(0, InboundDelay - 9.500);
 BF2 = max(0, 9.500 - InboundDelay );
 BF3 = max(0, TurnaroundTime - 82.250) * BF1;
 BF4 = max(0, 82.250 - TurnaroundTime ) * BF1;
 BF5 = max(0, GDPHoldingTime - 50.000);
 BF6 = max(0, 50.000 - GDPHoldingTime );
 BF7 = max(0, CarrierDelay - 0.600) * BF6;
 BF8 = max(0, 0.600 - CarrierDelay ) * BF6;
 BF9 = max(0, TurnaroundTime - 62.330) * BF6;
 BF10 = max(0, 62.330 - TurnaroundTime ) * BF6;
 BF12 = max(0, 72.000 - TurnaroundTime ) * BF2;
```

BF13 = max(0, CarrierDelay - 3.000) * BF2;
BF14 = max(0, 3.000 - CarrierDelay) * BF2;
BF15 = max(0, NumberSeats - 146.660) * BF1;
BF16 = max(0, 146.660 - NumberSeats) * BF1;
BF17 = max(0, TurnaroundTime - 51.000) * BF6;

sqrt(AbsorbedDelay) = min(0, -1.441 - 0.049 * BF1 + 0.091 * BF2 - .386774E-03 * BF3
+ 0.001 * BF4 + 0.004 * BF5 - 0.019 * BF6 + .356751E-03 * BF7
- 0.012 * BF8 + .699310E-03 * BF9 + .666024E-03 * BF10
- 0.003 * BF12 - 0.001 * BF13 + 0.010 * BF14
+ .964258E-04 * BF15 - .645073E-03 * BF16 - .637666E-03 * BF17)

Airport: MDW

Basis Functions
===============

BF1 = max(0, InboundDelay - 7.750);
BF2 = max(0, 7.750 - InboundDelay);
BF3 = max(0, TurnaroundTime - 89.000) * BF1;
BF4 = max(0, 89.000 - TurnaroundTime) * BF1;
BF5 = max(0, TurnaroundTime - 60.000);
BF6 = max(0, 60.000 - TurnaroundTime);
BF7 = max(0, CarrierDelay - 1.667);
BF8 = max(0, 1.667 - CarrierDelay);
BF9 = max(0, InboundDelay - 5.000) * BF6;
BF10 = max(0, 5.000 - InboundDelay) * BF6;
BF11 = max(0, GDPHoldingTime - 6.000);
BF12 = max(0, 6.000 - GDPHoldingTime);
BF13 = max(0, InboundDelay - 63.500) * BF8;
BF14 = max(0, 63.500 - InboundDelay) * BF8;
BF15 = max(0, TurnaroundTime - 227.000) * BF12;
BF16 = max(0, 227.000 - TurnaroundTime) * BF12;
BF17 = max(0, InboundDelay + 15.000) * BF8;

sqrt(AbsorbedDelay) = min(0, -2.156 - 0.042 * BF1 + 0.079 * BF2 - .109691E-03 * BF3
+ 0.001 * BF4 + 0.013 * BF5 + 0.041 * BF6 + 0.008 * BF7
+ 3.267 * BF8 - .537232E-03 * BF9 - 0.002 * BF10
+ 0.008 * BF11 - 0.301 * BF12 + 0.061 * BF13
- 0.042 * BF14 - 0.003 * BF15 + 0.001 * BF16
- 0.052 * BF17)

Airport: MEM

Basis Functions
===============

BF1 = max(0, InboundDelay - 13.250);
BF2 = max(0, 13.250 - InboundDelay);
BF3 = max(0, GDPHoldingTime - 80.000);
BF4 = max(0, 80.000 - GDPHoldingTime);
BF5 = max(0, TurnaroundTime - 88.000) * BF1;
BF6 = max(0, 88.000 - TurnaroundTime) * BF1;

BF7 = max(0, TurnaroundTime - 33.000);
BF8 = max(0, 33.000 - TurnaroundTime);
BF9 = max(0, CarrierDelay - 5.000);
BF10 = max(0, 5.000 - CarrierDelay);
BF11 = max(0, InboundDelay + 9.000) * BF4;
BF12 = max(0, - 9.000 - InboundDelay) * BF4;
BF13 = max(0, InboundDelay - 9.000) * BF4;
BF15 = max(0, NumberSeats - 134.600);
BF16 = max(0, 134.600 - NumberSeats);
BF17 = max(0, InboundDelay - 73.000) * BF7;
BF18 = max(0, 73.000 - InboundDelay) * BF7;

sqrt(AbsorbedDelay) = min(0, 0.043 - 0.058 * BF1 + 0.007 * BF2 - 0.005 * BF3 - 0.004 * BF4
+ .325583E-03 * BF5 + .972446E-03 * BF6 - 0.006 * BF7
+ 0.097 * BF8 + 0.004 * BF9 - 0.110 * BF10 - 0.001 * BF11
- .293752E-04 * BF12 + 0.001 * BF13 + 0.017 * BF15
+ 0.002 * BF16 + 0.002 * BF17 + .627970E-04 * BF18)

Airport: MIA

Basis Functions
===============

BF1 = max(0, InboundDelay - 12.600);
BF2 = max(0, 12.600 - InboundDelay);
BF3 = max(0, TurnaroundTime - 110.000) * BF1;
BF4 = max(0, 110.000 - TurnaroundTime) * BF1;
BF5 = max(0, GDPHoldingTime - 58.500);
BF6 = max(0, 58.500 - GDPHoldingTime);
BF7 = max(0, CarrierDelay - 11.000) * BF6;
BF8 = max(0, 11.000 - CarrierDelay) * BF6;
BF9 = max(0, NumberSeats - 168.330) * BF1;
BF10 = max(0, 168.330 - NumberSeats) * BF1;
BF11 = max(0, InboundDelay + 11.000);
BF13 = max(0, TurnaroundTime - 67.000) * BF11;
BF14 = max(0, 67.000 - TurnaroundTime) * BF11;
BF15 = max(0, NumberSeats - 145.000) * BF11;
BF16 = max(0, 145.000 - NumberSeats) * BF11;
BF17 = max(0, InboundDelay - 11.250) * BF6;
BF18 = max(0, 11.250 - InboundDelay) * BF6;

sqrt(AbsorbedDelay) = min(0, 1.340 + 0.062 * BF1 - 0.065 * BF2 - .865263E-04 * BF3
+ .531585E-03 * BF4 + 0.009 * BF5 - 0.018 * BF6
+ .936629E-04 * BF7 - .949757E-03 * BF8 - .468192E-03 * BF9
- .499283E-03 * BF10 - 0.115 * BF11 - .160857E-04 * BF13
+ 0.001 * BF14 + .351647E-03 * BF15 - .382989E-04 * BF16
+ .320360E-03 * BF17 + 0.001 * BF18)

Airport: MSP

Basis Functions
===============

BF1 = max(0, InboundDelay - 26.500);
BF2 = max(0, 26.500 - InboundDelay);
BF4 = max(0, 185.000 - TurnaroundTime) * BF1;
BF6 = max(0, 100.000 - CarrierDelay);
BF8 = max(0, 99.250 - TurnaroundTime) * BF6;
BF10 = max(0, 102.800 - TurnaroundTime) * BF2;
BF12 = max(0, 96.000 - GDPHoldingTime) * BF6;
BF13 = max(0, InboundDelay - 9.333) * BF6;
BF14 = max(0, 9.333 - InboundDelay) * BF6;
BF15 = max(0, InboundDelay + 14.000);
BF17 = max(0, TurnaroundTime - 94.000) * BF15;
BF18 = max(0, 94.000 - TurnaroundTime) * BF15;

sqrt(AbsorbedDelay) = min(0, 4.098 + 0.072 * BF1 - 0.074 * BF2 + .362963E-03 * BF4
- 0.018 * BF6 + .165202E-03 * BF8 - .524350E-03 * BF10
- .149700E-03 * BF12 - .407269E-03 * BF13 + .923777E-03 * BF14
- 0.116 * BF15 + .629995E-04 * BF17 + .566784E-03 * BF18)

Airport: ORD

Basis Functions
===============

BF1 = max(0, InboundDelay - 10.667);
BF2 = max(0, 10.667 - InboundDelay);
BF4 = max(0, 65.080 - TurnaroundTime) * BF1;
BF5 = max(0, CarrierDelay - 4.714);
BF6 = max(0, 4.714 - CarrierDelay);
BF7 = max(0, TurnaroundTime - 70.500);
BF8 = max(0, 70.500 - TurnaroundTime);
BF10 = max(0, 200.000 - NumberSeats);
BF11 = max(0, GDPHoldingTime - .142701E-06);
BF13 = max(0, 69.250 - TurnaroundTime) * BF2;
BF14 = max(0, InboundDelay - 60.500) * BF7;
BF15 = max(0, 60.500 - InboundDelay) * BF7;
BF17 = max(0, 33.500 - InboundDelay) * BF10;

sqrt(AbsorbedDelay) = min(0, -2.194 - 0.012 * BF1 + 0.078 * BF2 + .754285E-03 * BF4
+ 0.015 * BF5 - 0.057 * BF6 + 0.012 * BF7 + 0.035 * BF8
- 0.013 * BF10 + 0.009 * BF11 - 0.002 * BF13
- .311570E-03 * BF14 - .160500E-03 * BF15 + .262721E-03 * BF17)

Airport: PDX

Basis Functions
===============

BF1 = max(0, InboundDelay - 15.000);
BF2 = max(0, 15.000 - InboundDelay);
BF4 = max(0, 85.000 - TurnaroundTime) * BF1;
BF5 = max(0, CarrierDelay - 2.000);
BF6 = max(0, 2.000 - CarrierDelay);
BF7 = max(0, TurnaroundTime - 52.600);
BF8 = max(0, 52.600 - TurnaroundTime);
BF9 = max(0, InboundDelay - 16.000) * BF8;
BF10 = max(0, 16.000 - InboundDelay) * BF8;
BF12 = max(0, - 9.500 - InboundDelay) * BF6;
BF13 = max(0, CarrierDelay - 2.000) * BF2;
BF14 = max(0, 2.000 - CarrierDelay) * BF2;
BF15 = max(0, GDPHoldingTime + .205833E-07) * BF6;
BF17 = max(0, 4.000 - LegNumber) * BF2;
BF18 = max(0, NumberSeats - 30.000) * BF1;

sqrt(AbsorbedDelay) = min(0, -2.472 - 0.079 * BF1 + 0.100 * BF2 + 0.001 * BF4 + 0.017 * BF5
- 0.575 * BF6 - 0.004 * BF7 + 0.084 * BF8 - 0.001 * BF9
- 0.004 * BF10 - 0.086 * BF12 - .655613E-03 * BF13
+ 0.027 * BF14 + 0.007 * BF15 - 0.007 * BF17
+ .238990E-03 * BF18)

Airport: PHL

Basis Functions
===============

BF1 = max(0, InboundDelay - 10.000);
BF2 = max(0, 10.000 - InboundDelay);
BF3 = max(0, TurnaroundTime - 128.750) * BF1;
BF4 = max(0, 128.750 - TurnaroundTime) * BF1;
BF5 = max(0, CarrierDelay - 6.000);
BF6 = max(0, 6.000 - CarrierDelay);
BF7 = max(0, ScheduleDepartureTime - 71.000);
BF8 = max(0, 71.000 - ScheduleDepartureTime);
BF10 = max(0, 54.000 - InboundDelay) * BF6;
BF12 = max(0, 86.000 - TurnaroundTime);
BF13 = max(0, NumberSeats - 70.000);
BF14 = max(0, 70.000 - NumberSeats);
BF15 = max(0, GDPHoldingTime + .557923E-07) * BF6;
BF17 = max(0, 11.000 - InboundDelay) * BF12;
BF18 = max(0, Weight - 4.000);

sqrt(AbsorbedDelay) = min(0, -2.573 - 0.036 * BF1 + 0.092 * BF2 + .190306E-03 * BF3
+ .417888E-03 * BF4 + 0.016 * BF5 - 0.247 * BF6
- 0.013 * BF7 - 0.011 * BF8 + 0.004 * BF10
+ 0.027 * BF12 + 0.018 * BF13 - 0.044 * BF14
+ 0.002 * BF15 - 0.001 * BF17 - 0.899 * BF18)

Airport: PHX

Basis Functions
===============

BF1 = max(0, InboundDelay - 8.000);
BF2 = max(0, 8.000 - InboundDelay);
BF3 = max(0, TurnaroundTime - 134.660) * BF1;
BF4 = max(0, 134.660 - TurnaroundTime) * BF1;
BF5 = max(0, TurnaroundTime - 77.000);
BF6 = max(0, 77.000 - TurnaroundTime);
BF7 = max(0, CarrierDelay - 0.286);
BF8 = max(0, 0.286 - CarrierDelay);
BF9 = max(0, InboundDelay - 46.000) * BF8;
BF10 = max(0, 46.000 - InboundDelay) * BF8;
BF11 = max(0, TurnaroundTime - 80.000) * BF2;
BF12 = max(0, 80.000 - TurnaroundTime) * BF2;
BF13 = max(0, InboundDelay + 11.500);
BF15 = max(0, LegNumber - 2.000) * BF6;
BF16 = max(0, 2.000 - LegNumber) * BF6;
BF17 = max(0, LegNumber - 2.200);
BF18 = max(0, 2.200 - LegNumber);

sqrt(AbsorbedDelay) = min(0, -1.094 + 0.016 * BF2 - .883280E-04 * BF3 + .849202E-03 * BF4
+ 0.008 * BF5 + 0.040 * BF6 + 0.009 * BF7 - 5.018 * BF8
+ 0.126 * BF9 + 0.094 * BF10 - .517526E-03 * BF11
- 0.001 * BF12 - 0.090 * BF13 - 0.002 * BF15
- 0.040 * BF16 + 0.087 * BF17 + 0.706 * BF18)

Airport: PIT

Basis Functions
===============

BF1 = max(0, InboundDelay - 13.500);
BF2 = max(0, 13.500 - InboundDelay);
BF3 = max(0, GDPHoldingTime - 5.000);
BF4 = max(0, 5.000 - GDPHoldingTime);
BF5 = max(0, TurnaroundTime - 104.500) * BF1;
BF6 = max(0, 104.500 - TurnaroundTime) * BF1;
BF7 = max(0, CarrierDelay - 1.500) * BF4;
BF8 = max(0, 1.500 - CarrierDelay) * BF4;
BF9 = max(0, NumberSeats - 79.000);
BF10 = max(0, 79.000 - NumberSeats);
BF11 = max(0, InboundDelay + 10.500);
BF13 = max(0, TurnaroundTime - 104.500) * BF11;
BF14 = max(0, 104.500 - TurnaroundTime) * BF11;
BF15 = max(0, ScheduleEnrouteWeather - 78.500) * BF11;
BF16 = max(0, 78.500 - ScheduleEnrouteWeather) * BF11;
BF17 = max(0, Distance - 287.500) * BF2;
BF18 = max(0, 287.500 - Distance) * BF2;

sqrt(AbsorbedDelay) = min(0, -0.272 + 0.143 * BF1 - 0.023 * BF2 + 0.010 * BF3 + 0.013 * BF4

- 0.002 * BF5 - .946825E-03 * BF6 + 0.002 * BF7
- 0.072 * BF8 + 0.010 * BF9 + 0.009 * BF10
- 0.169 * BF11 + 0.001 * BF13 + 0.001 * BF14
+ .218123E-05 * BF15 - .904125E-04 * BF16 - .803026E-05 * BF17
+ .283192E-03 * BF18)

Airport: SAN

Basis Functions
===============

BF1 = max(0, InboundDelay - 13.667);
BF2 = max(0, 13.667 - InboundDelay);
BF4 = max(0, 1.000 - CarrierDelay);
BF5 = max(0, InboundDelay + 8.333) * BF4;
BF6 = max(0, - 8.333 - InboundDelay) * BF4;
BF7 = max(0, TurnaroundTime - 10.000) * BF1;
BF8 = max(0, NumberSeats - 130.000);
BF9 = max(0, 130.000 - NumberSeats);
BF10 = max(0, TurnaroundTime - 80.000);
BF11 = max(0, 80.000 - TurnaroundTime);
BF12 = max(0, NumberSeats - 35.000) * BF2;
BF13 = max(0, 35.000 - NumberSeats) * BF2;
BF14 = max(0, InboundDelay - 5.500) * BF4;
BF16 = max(0, TurnaroundTime - 78.660) * BF2;
BF17 = max(0, 78.660 - TurnaroundTime) * BF2;
BF18 = max(0, DepartureRatio_15 - 0.110) * BF11;

sqrt(AbsorbedDelay) = min(0, -3.413 + 0.177 * BF2 - 0.057 * BF5 - 0.068 * BF6 - .242539E-03 * BF7
+ 0.012 * BF8 - 0.017 * BF9 + 0.025 * BF10
+ 0.055 * BF11 - .584270E-03 * BF12 + 0.009 * BF13
+ 0.053 * BF14 - 0.001 * BF16 - 0.002 * BF17
- 0.010 * BF18)

Airport: SEA

Basis Functions
===============

BF1 = max(0, InboundDelay - 17.000);
BF2 = max(0, 17.000 - InboundDelay);
BF3 = max(0, TurnaroundTime - 98.000) * BF1;
BF4 = max(0, 98.000 - TurnaroundTime) * BF1;
BF5 = max(0, TurnaroundTime - 95.000);
BF6 = max(0, 95.000 - TurnaroundTime);
BF7 = max(0, CarrierDelay - 0.500);
BF8 = max(0, 0.500 - CarrierDelay);
BF9 = max(0, InboundDelay - 48.500) * BF8;
BF10 = max(0, 48.500 - InboundDelay) * BF8;
BF11 = max(0, InboundDelay + 10.500) * BF6;
BF12 = max(0, - 10.500 - InboundDelay) * BF6;
BF13 = max(0, InboundDelay + 9.667);
BF15 = max(0, CarrierDelay + .455655E-07) * BF2;

```
 BF16 = max(0, NumberSeats - 93.330) * BF6;
 BF17 = max(0, 93.330 - NumberSeats ) * BF6;
 BF18 = max(0, GDPHoldingTime - .179148E-07);

sqrt(AbsorbedDelay) = min(0, 0.235 + 0.109 * BF1 - 0.030 * BF2 - .120898E-03 * BF3
          - .632901E-03 * BF4 + 0.003 * BF5 - 0.004 * BF6
          + 0.016 * BF7 - 3.268 * BF8 + 0.031 * BF9 + 0.060 * BF10
          + 0.001 * BF11 + .233143E-03 * BF12 - 0.149 * BF13
          - .554835E-03 * BF15 + .988322E-04 * BF16 + .146373E-03 * BF17
          + 0.011 * BF18)

Airport: SFO

 Basis Functions
 ===============

 BF1 = max(0, InboundDelay - 18.500);
 BF2 = max(0, 18.500 - InboundDelay );
 BF4 = max(0, 139.600 - TurnaroundTime ) * BF1;
 BF5 = max(0, CarrierDelay - 6.000);
 BF6 = max(0, 6.000 - CarrierDelay );
 BF8 = max(0, 64.000 - TurnaroundTime );
 BF10 = max(0, 65.420 - TurnaroundTime ) * BF2;
 BF11 = max(0, InboundDelay + 11.000) * BF6;
 BF12 = max(0, - 11.000 - InboundDelay ) * BF6;
 BF13 = max(0, InboundDelay - 41.500) * BF6;
 BF15 = max(0, GDPHoldingTime + .390500E-07) * BF6;
 BF16 = max(0, CarrierDelay - 0.500) * BF8;
 BF18 = max(0, ScheduleDepartureTime - 24.000) * BF8;

sqrt(AbsorbedDelay) = min(0, -2.787 - 0.050 * BF1 + 0.073 * BF2 + .566958E-03 * BF4
          + 0.012 * BF5 + 0.050 * BF8 - 0.002 * BF10
          - 0.006 * BF11 - 0.010 * BF12 + 0.006 * BF13
          + 0.004 * BF15 - .651956E-03 * BF16 + .401657E-03 * BF18)

Airport: SLC

 Basis Functions
 ===============

 BF1 = max(0, InboundDelay - 16.000);
 BF2 = max(0, 16.000 - InboundDelay );
 BF3 = max(0, TurnaroundTime - 73.110) * BF1;
 BF4 = max(0, 73.110 - TurnaroundTime ) * BF1;
 BF5 = max(0, TurnaroundTime - 78.500);
 BF6 = max(0, 78.500 - TurnaroundTime );
 BF7 = max(0, TurnaroundTime - 80.000) * BF2;
 BF8 = max(0, 80.000 - TurnaroundTime ) * BF2;
 BF9 = max(0, CarrierDelay - 0.333);
 BF10 = max(0, 0.333 - CarrierDelay );
 BF12 = max(0, - 10.000 - InboundDelay ) * BF10;
 BF13 = max(0, Weight - 4.570);
 BF14 = max(0, 4.570 - Weight );
```

BF15 = max(0, CarrierDelay - 2.800) * BF2;
BF16 = max(0, 2.800 - CarrierDelay) * BF2;
BF17 = max(0, GDPHoldingTime - 70.000) * BF10;
BF18 = max(0, 70.000 - GDPHoldingTime) * BF10;

sqrt(AbsorbedDelay) = min(0, -3.664 - 0.029 * BF1 + 0.109 * BF2 - .220649E-03 * BF3
+ .861071E-03 * BF4 + 0.011 * BF5 + 0.045 * BF6
- .506022E-03 * BF7 - 0.002 * BF8 + 0.022 * BF9
- 0.262 * BF12 + 0.260 * BF13 + 0.772 * BF14
- .822847E-03 * BF15 + 0.012 * BF16 + 0.082 * BF17
- 0.028 * BF18)

Airport: STL

Basis Functions
===============

BF1 = max(0, InboundDelay - 6.667);
BF2 = max(0, 6.667 - InboundDelay);
BF3 = max(0, TurnaroundTime - 42.400);
BF4 = max(0, 42.400 - TurnaroundTime);
BF5 = max(0, ScheduleDepartureTime - 65.000);
BF6 = max(0, 65.000 - ScheduleDepartureTime);
BF8 = max(0, 37.000 - InboundDelay) * BF3;
BF9 = max(0, NumberSeats - 50.000);
BF10 = max(0, 50.000 - NumberSeats);
BF11 = max(0, InboundDelay + 7.500) * BF9;
BF12 = max(0, - 7.500 - InboundDelay) * BF9;
BF13 = max(0, TurnaroundTime - 99.500) * BF1;
BF14 = max(0, 99.500 - TurnaroundTime) * BF1;
BF16 = max(0, 49.600 - GDPHoldingTime);
BF18 = max(0, 0.600 - CarrierDelay);

sqrt(AbsorbedDelay) = min(0, -1.137 - 0.059 * BF1 + 0.087 * BF2 - 0.051 * BF3 + 0.028 * BF4
- 0.016 * BF5 - 0.013 * BF6 + 0.001 * BF8 + 0.004 * BF9
+ 0.068 * BF10 + .133845E-03 * BF11 - .885348E-03 * BF12
+ 0.003 * BF13 + .693897E-03 * BF14 - 0.014 * BF16
- 0.682 * BF18)

Airport: TPA

Basis Functions
===============

BF1 = max(0, InboundDelay - 6.000);
BF2 = max(0, 6.000 - InboundDelay);
BF3 = max(0, TurnaroundTime - 80.000) * BF1;
BF4 = max(0, 80.000 - TurnaroundTime) * BF1;
BF5 = max(0, GDPHoldingTime - 30.000);
BF6 = max(0, 30.000 - GDPHoldingTime);
BF8 = max(0, 56.500 - TurnaroundTime) * BF6;
BF10 = max(0, 80.000 - TurnaroundTime) * BF2;
BF11 = max(0, InboundDelay - 13.000) * BF6;
BF12 = max(0, 13.000 - InboundDelay) * BF6;

BF13 = max(0, ScheduleDepartureTime - 74.000) * BF6;
BF14 = max(0, 74.000 - ScheduleDepartureTime) * BF6;
BF16 = max(0, 5.000 - CarrierDelay) * BF6;
BF17 = max(0, Distance - 558.500);
BF18 = max(0, 558.500 - Distance);

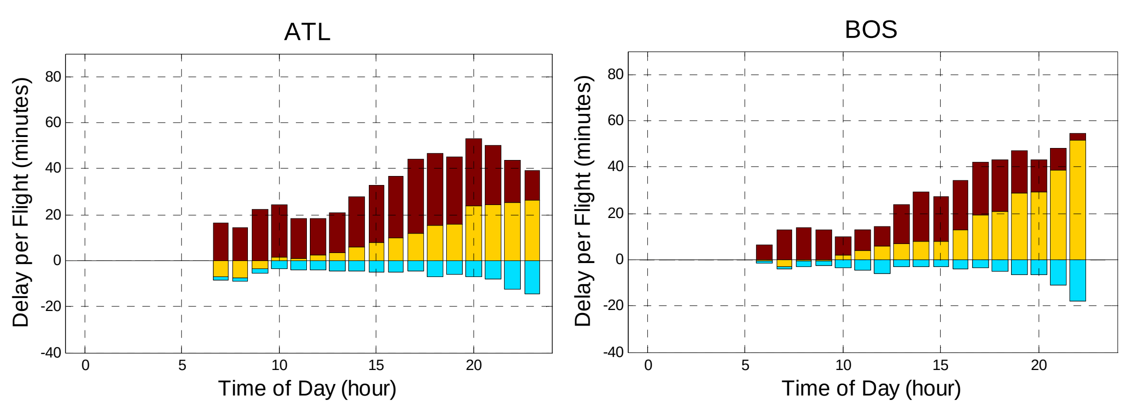
sqrt(AbsorbedDelay) = min(0, -1.075 - 0.042 * BF1 + 0.109 * BF2 - .181339E-03 * BF3
+ .634843E-03 * BF4 + 0.007 * BF5 - 0.037 * BF6
+ 0.002 * BF8 - 0.003 * BF10 + .721148E-03 * BF11
+ 0.002 * BF12 - 0.003 * BF13 - .444700E-03 * BF14
- 0.004 * BF16 + .266903E-03 * BF17 + 0.002 * BF18)

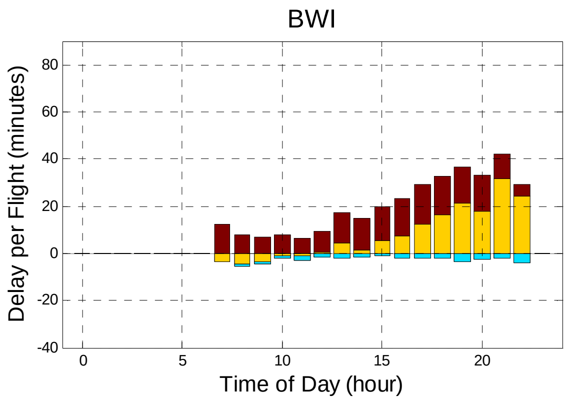
APPENDIX C

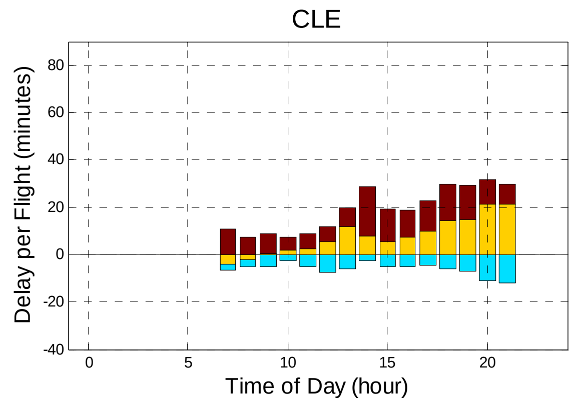
AGGREGATED DELAYS FOR A DAY IN 34 OEP AIRPORTS

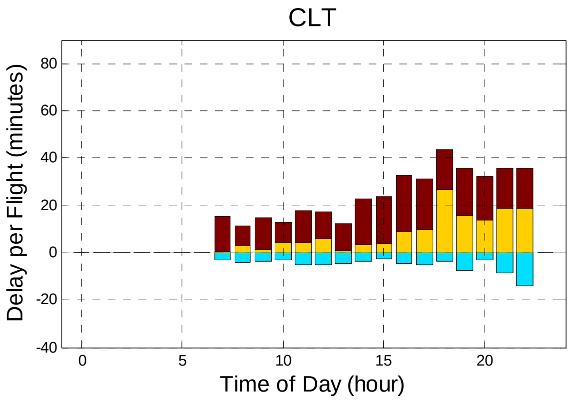
The figures in Figure **C.1** were drawn for connecting flights in June, July and August 2006. By calling them connecting flights, we mean their previous leg can be found in BTS database. For any specific hour, if the number of total flights in three months is less than 46, its aggregate delay is not shown on the figure. Each bar shows the components of the average wheels-off delay for each flight scheduled to depart in a given 60 minute period. The average delay experienced during this period is the inbound delay (lightest, yellow) plus the generated delay (darkest, dark red) minus the absorbed delay (medium dark, cyan).

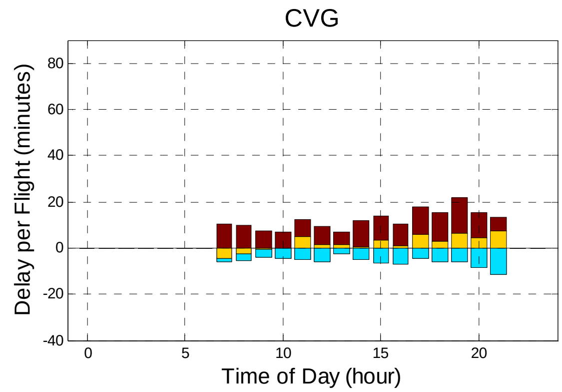
Figure **C.1**: Aggregated Inbound Delay, Airport Generated Delay, and Airport Absorbed Delay for each hour in Summer 2006 at OEP 34 Airports

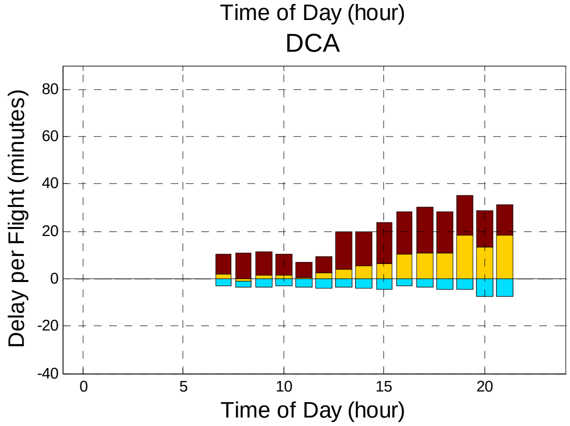


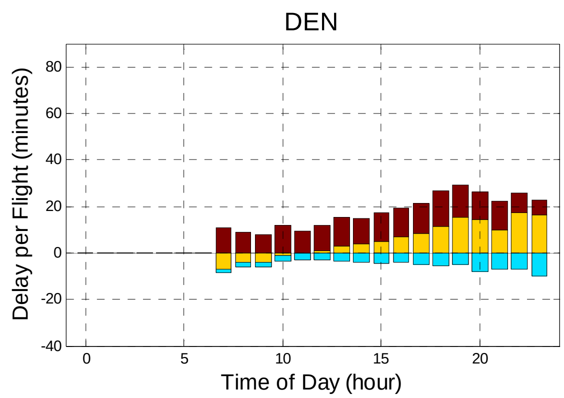

BWI
Delay per Flight (minutes)
Time of Day (hour)
80
60
40
20
0
-20
-40
0
5
10
15
20

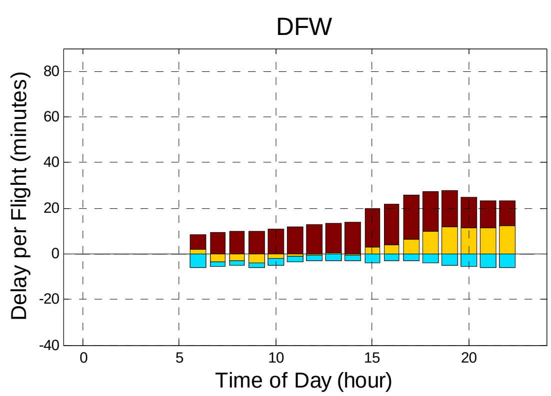

CLE
Delay per Flight (minutes)
Time of Day (hour)
80
60
40
20
0
-20
-40
0
5
10
15
20

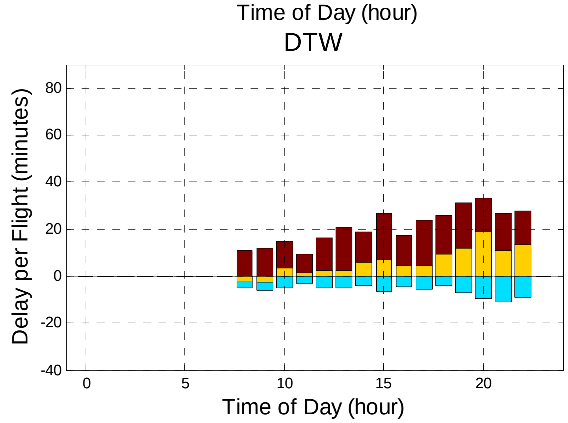

CLT
Delay per Flight (minutes)
Time of Day (hour)
80
60
40
20
0
-20
-40
0
5
10
15
20

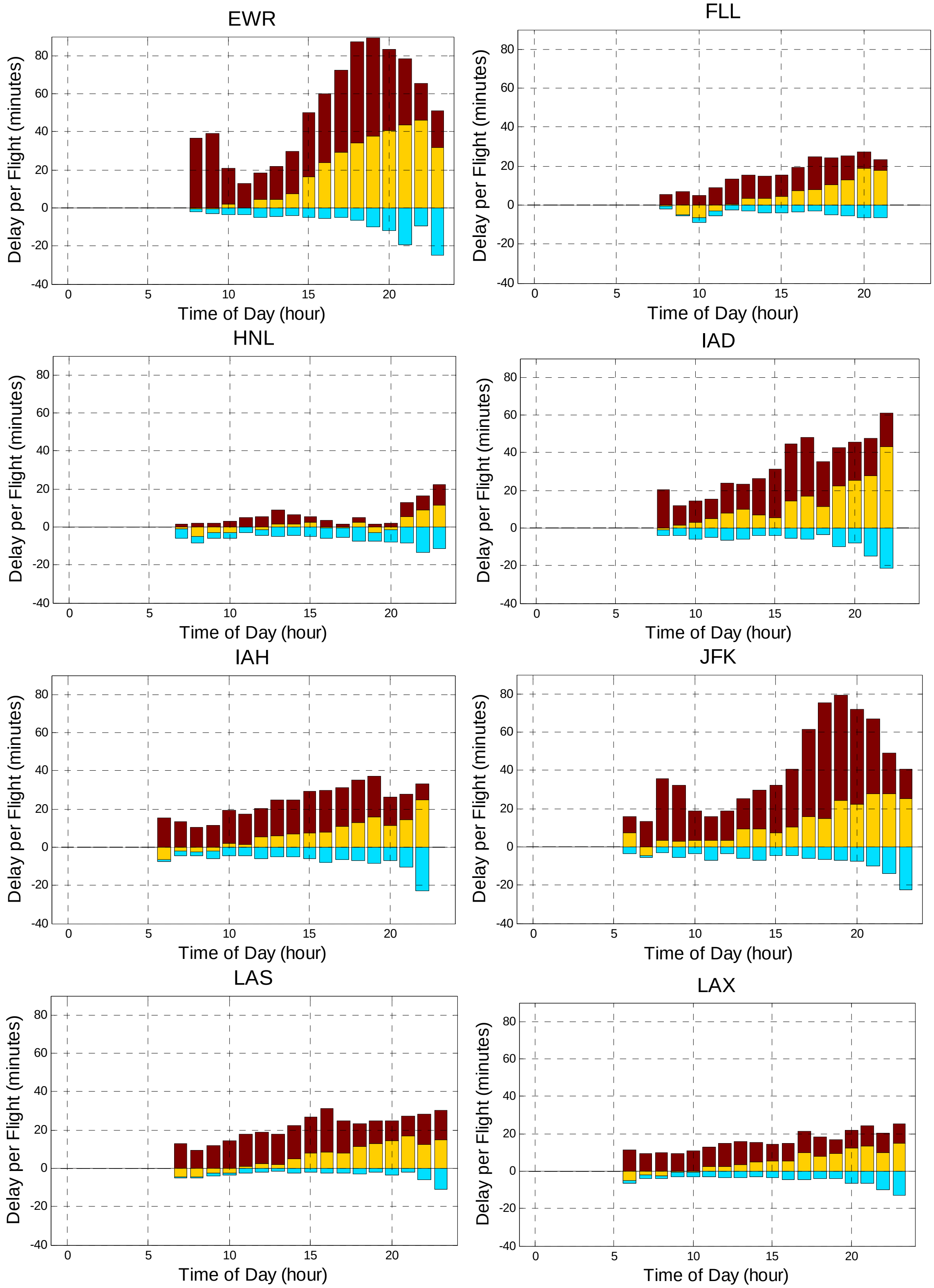

CVG
Delay per Flight (minutes)
Time of Day (hour)
80
60
40
20
0
-20
-40
0
5
10
15
20

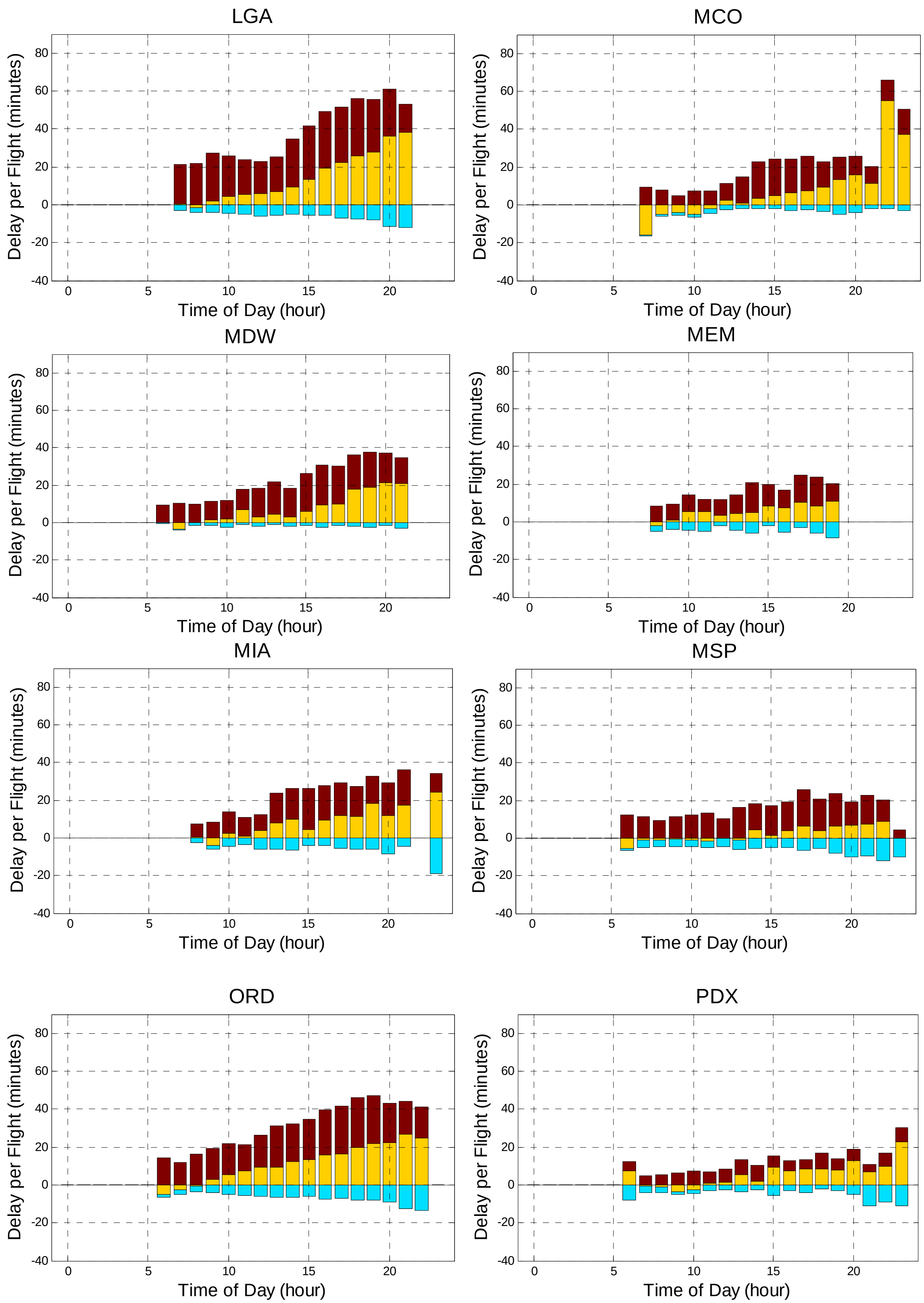

DCA
Delay per Flight (minutes)
Time of Day (hour)
80
60
40
20
0
-20
-40
0
5
10
15
20

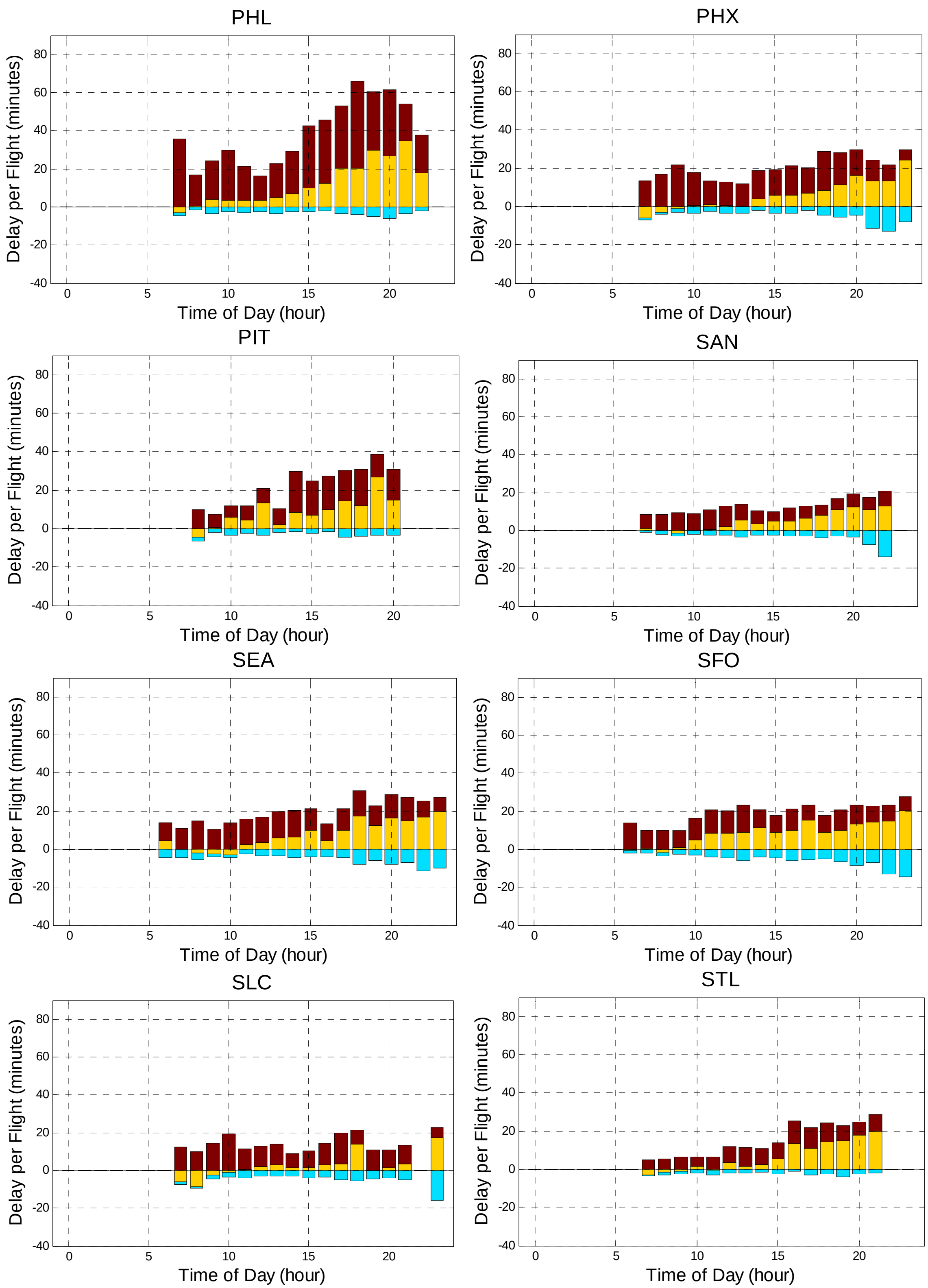

DEN
Delay per Flight (minutes)
Time of Day (hour)
80
60
40
20
0
-20
-40
0
5
10
15
20

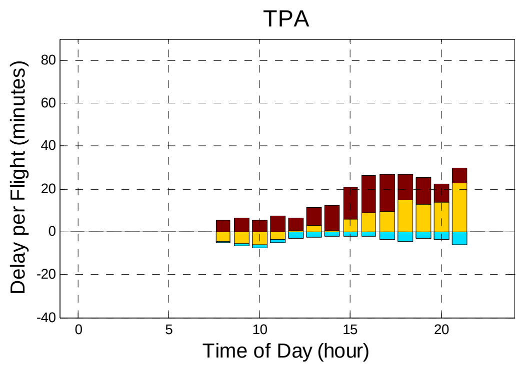

DFW
Delay per Flight (minutes)
Time of Day (hour)
80
60
40
20
0
-20
-40
0
5
10
15
20


DTW
Delay per Flight (minutes)
Time of Day (hour)
80
60
40
20
0
-20
-40
0
5
10
15
20


EWR
FLL
HNL
IAD
IAH
JFK
LAS
LAX
Delay per Flight (minutes)
Time of Day (hour)


LGA
MCO
MDW
MEM
MIA
MSP
ORD
PDX
Delay per Flight (minutes)
Time of Day (hour)
-40
-20
0
20
40
60
80
0
5
10
15
20


PHL
PHX
PIT
SAN
SEA
SFO
SLC
STL
Delay per Flight (minutes)
Time of Day (hour)
80
60
40
20
0
-20
-40
0
5
10
15
20


TPA
Delay per Flight (minutes)
80
60
40
20
0
-20
-40
0
5
10
15
20
Time of Day (hour)

BIBLIOGRAPHY

Abdel-Aty, M.A., C. Lee, Y. Bai, X. Li, and M. Michalak (2007). Two-Stage Approach to Identify Flight Delay Patterns. TRB 86th Annual Meeting Compendium of Papers CD-ROM.

Allan, S. S., J. A. Beesley, J. E. Evans and S. G. Gaddy (2001). Analysis of Delay Causality at Newark International Airport. Proceedings of 4th USA/Europe Air Traffic Management R\&D Seminar, Santa Fe, New Mexico, 3-7 December 2001

Allan,S. S., R. DeLaura, B. Martin, D.A. Clark and C. Gross (2004). Advanced Terminal Weather Products Demonstration in New York. Proceedings of 11th Conference on Aviation, Range and Aerospace Meteorology, Hyannis, MA 2004

Allan, S.S., S. Gaddy and J. Evans (2001). Delay Causality and Reduction at the New York City Airports Using Terminal Weather Information Systems. Massachusetts Institute of Technology,Project report, ATC-291, 16 February 2001

Atkins, S. and D. Walton (2002). Prediction and Control of Departure Runway Balancing at Dallas/Fort Worth Airport. Proceedings of American Control Conference Anchorage, AK. May 2002.

Baden, W., J. DeArmon, J. Kee and L Smith (2005). Assessing Schedule Delay Propagation in the National Airspace System. Proceedings of 47th Annual Transportation Research Forum.

Baden,W., J. DeArmon and J. Kee (2006). Assessing Airports Delay Influence on the National Airspace System, MITRE Library

Ball, M. O. and G. Lulli (2004). Ground Delay Programs: Optimizing over the Included flight Set Based on Distance. Air Traffic Control Quarterly, 12, 1-25.

Ball, M. O., A. Mukherjee, and B. Subramanian (2006). Models for Estimating Monthly Delays and Cancellations in the NAS, Nextor presentation. 12 December 2006. http://www.nextor.org/NAS2006/Session%203a_Mukherjee_Modeling%20Delays%20and%20Cancellations.pdf Accessed July 16,2007

Beatty, R., R. Hsu, L. Berry, J. Rome, (1999). Preliminary Evaluation of Flight Delay Propagation Through an Airline Schedule. Air Traffic Control Quarterly, 7, 1999, pp. 259-270.

Bilimoria, K. and B. Sridhar. (2000). FACETL Future ATM Concepts Evaluation Tool. In Proceedings of 3rd USA/Europe Air Traffic Management R&D Seminar, Napoli, Italy, 2000.

Boswell, S. and J. Evans (1997). Analysis of Downstream Impacts of Air Traffic Delay. Project Report ATC-257, MIT Lincoln Laboratory, Cambridge, MA.

Bhadra, D. and P. Texter (2005), Airline Networks An Econometric Framework to Analyze Domestic US Air Travel, Journal of Transportation and Statistics, Vol. 7 Number 1.

Brockwell, P. J. and R. A. Davis (1996). Introduction to Time Series and Forecasting. Springer.

(BTS 1) BTS sources: available at http://www.bts.gov/programs/airline_information/sources/

(BTS 2) BTS Airline On-Time Performance Data table profile: available at http://www.transtats.bts.gov/TableInfo.asp?Table_ID=236&DB_Short_Name=On-Time&Info_Only=0

(BTS 3) BTS delay cause definition: available at http://www.bts.gov/help/aviation.html#q3

(BTS 4) BTS glossary: available at http://www.transtats.bts.gov/Glossary.asp?index=C.

Callaham, M.B., J. S. DeArmon, A. M. Copper, J. H. Goodfriend, D. Monch-Mooney, and G. H. Solomos (2001). Assessing NAS Performance: Normalizing for the Effects of Weather. Proceedings of 4th USA/Europe Air Traffic Management R\&D Symposium, Santa Fe, NM. 3-7 December, 2001

Chatterji, G. B., B. Sridhar, K. Sheth, D. Kim, and D. Mulfinger (2004). Methods for Establishing Confidence Bounds on Sector Demand Forecasts. Proceedings of AIAA Guidance, Navigation, and Control Conference, Providence, RI, August 16-19, 2004.

Chatterji, G. B. and B. Sridhar (2005). National airspace System Delay Estimation Using Weather Weighted Traffic Counts. AIAA Guidance, Navigation, and Control conference and Exhibit, San Francisco, California

Churchill, A. M., D. J. Lovell and M. O. Ball (2007). Examining the Temporal Evolution of Propagated Delays at Individual Airports: Case Studies. Proceedings of 7th USA\/Europe Air Traffic Management R\&D Seminar, Barcelona, Spain, 7,2007

Dai, D. M. (2006). Delay Prediction Models for Departure Flights, TRB 2006 Annual Meeting CD-ROM.

Danks D. Summary of the Bayes Net Formalism. Available at www.jsmf.org/meetings/2003/danks-summary.ppt

DeArmon, J. S. (1992). Analysis and research for traffic flow management. Proceedings of 37th Annual Conference of Air Traffic Control Association, Atlantic City, NJ, Air Traffic Control Association, 423-429.

DeArmon, J. S. (1993). Do air traffic flow problems interact? A preliminary analysis. Proceedings of IEEE Transactions on Control Systems Technology, 1(3). pp 195-203.

Donohue, G. L. (2007) Transforming the National Airspace System. Presentation at Integrated Communications Navigation and Surveillance (ICNS) Conference, May 20 2003. http://spacecom.grc.nasa.gov/icnsconf/docs/2003/01_Plenary/PS-09-Donohue.pdf. Accessed July 2007.

Evans, J. E., S. Allan and M. Robinson (2004). Quantifying Delay Reduction Benefits for Aviation Convective Weather Decision Support Systems. Proceedings of the 11th Conference on Aviation. Range and Aerospace Meteorology, Hyannis, MA

Evans, A.D. and J. P. Clark (2005). Responses to Airport Delays – A System Study of Newark International Airport. http://icat-server.mit.edu/library/fullRecord.cgi?idDoc=117 Accessed July 16, 2005.

Fan, T. P. C., and A. R. Odoni (2001). The Potential of Demand Management as a Short-Term Means of Relieving Airport Congestion. Proceedings of the 6th USA/Europe Air Traffic Management R&D Seminar

Freund R. J., W. J. Wilson and P. Sa (2006). Regression Analysis: Statistical Modeling of a Response Variable. Academic Press.
Harback, K.T., and J. I. Daniel (2005). Do Airlines that Dominate Traffic at Hub Airports Experience Less Delay? Working Paper No. 2005-09, Department of Economics, Alfred Lerner College of Business & Economics University of Delaware.

Friedman N., M. Linial, I. Nachman, and D. Pe'er (2000). Using Bayesian Networks o Analyze Expression Data. Journal of Computational Biology. Vol. 7, Numbers 3.4, 2000, Mary Ann Lieert, Inc. Pp. 601-620.

Hamilton, J. D. (1994). Time Series Analysis. Princeton University Press.

Hansen, M. and T. Bolic (2001). Delay and Flight Time Normalization Procedures for Major Airports: LAX Case Study. Research Report. UCB-ITS-RR-2001-5. 2001

Hansen, M. and C.Y. Hsiao (2005). Going South? An Econometric Analysis of US Airline Flight Delays from 2000 to 2004. Proceedings of TRB 84th Annual Meeting, Washington, D.C.

Hansen, M. and C.Y. Hsiao (2006). An Econometric Analysis of U.S. Airline Flight Delays with Time-of-Day Effects. Journal of the Transportation Research Board

Hansen, M. and Y. Zhang (2004) Operational Consequences of Alternative Airport Demand Management Policies: The Case of LaGuardia Airport, In Journal of Transportation Research Record. 2004. Vol. 1888, pp. 15-21

Hafizogullari, S., P. Chinnusamy, and C. Tunasar (2002). Simulation Reduces Airline Misconnections: A Case Stud. Proceedings of the 2002 Winter Simulation conference.

Hemm, R. and G. Shapiro (1998). Terminal Area Productivity Airport Wind Analysis and Chicago O'Hare Model Description. NASA technical documentation. NASA/CR-1998-207662

Hoffman J. (2001). Demand dependence of throughput and delay at New York LaGuardia Airport. The MITRE Corporation 2001 Technical Paper. Available at http://www.mitre.org/work/tech_papers/tech_papers_01/hoffman_airport/

Penny, S., R. Hoffman, J. Krozel, and A. Roy (2005). Classification of Days in the National Airspace System Using Cluster Analysis. Air Traffic Control Quarterly, 13, 25–43.

Hoffman, J. (2007) Demand dependence of throughput and delay at New York LaGuardia Airport, The MITRE Corporation. http://www.mitre.org/work/tech_papers/tech_papers_01/hoffman_airport/hoffman.pdf Accessed July 16, 2007.

Hsiao, C.Y and M. Hansen (2006) Econometric Analysis of U.S. Airline Flight Delays with Time-of-Day Effects. Transportation Research Record, Volumn 1921/2006, Page 104-112 AIAA-2003-5711, Proceedings of AIAA Guidance, Navigation, and Control Conference, Austin, TX, August 11-14, 2003.

Idris, H., J. P. Clarke, R. Bhuva, and L. Kang (2002). Queuing Model for Taxi-Out Time Estimation. Air Traffic Control Quarterly. 10(1) 2002, pp. 1-22

Janic, M (2005). Modeling the Large Scale Disruptions of an Airline Network. Journal of Transportation Engineering 131, 249

Johnson, R. A. and D. W. Wichern (2002). Applied Multivariate Statistical Analysis, Fifth Edition, Prentice Hall.

Jones, J., lecture notes, available at http://www.richland.edu/james/lecture/m170/ch13-2wy.html

Krozel, J., B. Hoffman, S. Penny, and T. Butler (2003). Aggregate Statistics of the National Airspace System. Proceedings of the AIAA Guidance, Navigation and Control Conference.

Krozel, J., Bl. Capozzi, A. D. Andre, and P. Smith (2003). The Future National Airspace System: Design Requirements Imposed By Weather Constraints. AIAA-2003-5769, Proceedings of AIAA Guidance, Navigation, and Control Conference, Austin, TX.

Laskey, K. B. MEBN: A Logic for Open-World Probabilistic Reasoning. The Volnegau School of Information Technology and Engineering, George Mason University, available at http://ite.gmu.edu/~klaskey/index.html.

Laskey, K. B., N. Xu, and C.H. Chen (2006). Propagation of Delays in the National Airspace Syste. Proceedings of 22nd Conference on Uncertainty in Artificial Intelligence, Cambridge, MA.

Le, L., G. Donohue, and C.H. Chen (2003). Space-Time Correlation Analysis of Quality-of-Service at Major US Airports-O'Hare International and Minneapolis Airports: A Case Study. Proceedings of Digital Avionics Systems Conference, 2003. DASC'03. the 22nd Volume 2, pp. 10.A.5-10.1-8. http://ieeexplore.ieee.org/iel5/8816/27920/01245931.pdf Accessed July 16 2007.

Liou, J. S. (2006). Delay Prediction Models for Departure Flights. Proceedings of TRB 85th annual Meeting, Washington, D.C.

Long, D., V. Stouffer-Coston, P. Kostiuk, R. Kula, and B. Fernandez (2001). Integrating LMINET with TAAM and SIMMOD, A Feasibility Study. NASA-CR-2001-210875

MARS^{TM} User Guide. Salford Systems, 2001, pp. 36-37

Mueller, E. R. and G. B. Chatterji (2002). Analysis of Aircraft Arrival and Departure Delay Characteristics. AIAA 2002-5866. IAA's Aircraft Technology, Integration, and Operations (ATIO) Technical 1-3 October 2002, Los Angeles, California

Mukherjee, A., D. J. Lovell, M. O. Ball, A. R. Odoni, and G. Zerbib (2005). Modeling Delays and Cancellation Probabilities to Support Strategic Simulations. Proceedings of the 6th USA/Europe Air Traffic Management R&D Symposium, Baltimore, MD.

Northcott, R. (2006). Causal Efficacy and the Analysis of Variance, Journal of Biology and Philosophy (2006) 21:253-276

Odoni, A.R., J. Bowman, D. Delahaye, J.J. Deyst, E. Feron, R.J. Hansman, K. Khan, J.K. Kuchar, N. Pujet and R.W. Simpson (1997). Existing and Required Modeling Capability for Evaluating ATM Systems and Concepts. Final Report of Modeling Research Under

NASA/AATT. International Center for Air Transportation, Massachusetts Institute of Technology.

Odoni, A. R. (2004). Airport Characteristics. Airport Systems Course, Massachusetts Institute of Technology. Available at http://ardent.mit.edu/airports/ASP_current_lectures/Airport_Characteristics_04_bw2.pdf

Pearl, J. (2000) Causality: Models, Reasoning, and Inference. Cambridge University Press 2000.

Pearl, J. (1988) Probabilistic Reasoning in Intelligent Systems. Morgan and Kaufman, San Mateo.

Pepper, J. W., K. R. Mills and L A. Wojcik (2003). Predictability and Uncertainty in Air Traffic Flow Management. Proceedings of the 6th USA/Europe Air Traffic Management R\&D Seminar

Post, J., M. Bennett, J. Bonn and D. Knor. Estimation of an En Route Weather Severity Index Using Lightning Strike and Flight Plan Data. FAA Technical Publication, Available at http://ffp1.faa.gov/approach/media/estimation/estimation.htm

Post, J., J. Bonn, M. Bennett, D. Howell and D. Knor (2002) The Use of Flight Track and Convective Weather Densities for National airspace System Efficiency Analysis. IEEE/AIAA digital Aviation Systems conference (DASC).

DOT (2001). Reporting the Causes of Airline Delays and Cancellations. Department of Transportation, Office of the Secretary, 14 CFR Part 234 [Docket No. OST 2000-8164], RIN 2139-AA09

Rupp, N. G. (2005). Flight Delays and Cancellations. Working paper

U.S. Department of Transportation, "FAA Aerospace Forecasts Fiscal Years 2006-2017," Office of Aviation Policy and Plans, Tech. Rep., March 2005

Schaefer, L. and D. Milner (2001). Flight Delay Propagation Analysis with the Detailed Policy Assessment Tool. Proceedings of the 2001 IEEE Systems, Man and Cybernetics Conference

Shumway, R., H. and D. S. Stoffer (2000). Time Series Analysis and Its Applications. Springer.

Sinha, A. N. (2001). Weather and its Role in Aviation Delays and Safety: Anatomy of a Weather Delay. Proceedings of Aviation Gridlock: Understanding the Options and Seeking Solutions. Phase III: Weather and Weather Technology, Washington, D.C.

Steyerberg EW, MJ Eijkemans, FE Harrell, Jr., JD. Habbema (2000) Prognostic modeling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. Stat Med. 2000;19(8):1059-79.

StatSoft Electronic Textbook, available at http://www.statsoft.com/textbook/stcluan.html
Steinberg, D., B. Bernstein, P. Colla and K. Martin (2001).

MARS User Guide. San Diego, CA: Salford Systems, 2001.

Stergiou, C., and D. Siganos Neural Networks, SURPRISE 96 Journal Vol. 4

Sussman J. M. (2004) Lecture Note on 1.221J/11.527J/ESD.201J Transportation Systems.

Sussman, J. M. (2000) Introduction to Transportation Systems. Artech House Publishers, Boston and London.

Turner, S. M., and R.A. Mundy (1996). Hubs versus Hub-Nots: A Comparison of Various U.S. Airports. Journal of Air Transportation World Wide Vol.1, No.1

Tan, P.N, M. Steinbach, and V. Kumar (2005). Introduction to Data Mining. Addison-Wesley April 2005.

Vigneau, W. (2003) Flight Delay Propagation: Synthesis of the Study. EEC Note No. 18/03, Eurocontrol Agency, Brussels, Belgium.

Wang, P. T. R., N. Tene, and L. A. Wojcik (2002). Relationship Between Airport Congestion and At-Gate Delay. Proceedings of 21st DASC, pp.2.D.5.1 to 2.D.5.11.

Wang, P. T. R., L. A. Schaefer and L. A. Wojcik (2003). Flight Connections and Their Impacts on Delay Propagation. Proceedings of Digital Avionics Systems Conference, DASC'03.

Wang, D.Y. and L. Sherry (2006). Trend Analysis of Airline Passenger Trip Delays. Transportation Research Board, August 2006.

Welch, J. D., and S. Ahmed (2003). Spectral Analysis of Airport Capacity and Delay. Proceedings of the 5th USA/Europe Air Traffic Management R&D Symposium.

Willemain, T. R. (2001). Contingencies and Cancellations in Ground Delay Programs. NEXTOR report wp-01-3.

Wojcik (2001). DPAT: A Tool for System-Level Simulation of Airports and Airspace. Mitre DPAT Training Briefing.

Xu, N., K.B. Laskey, C.H. Chen, S.C. Williams and L. Sherry(2007). Bayesian Network Analysis of Flight Delays. Proceedings of Transportation Research Board 86th Annual Meeting Compendium of Papers CD-ROM, 2007.

Xu, Ning, (2003) A Comparison of Discretization Methods for Bayesian Networks. Technical Report and Presentation, Systems Engineering and Research Department, George Mason University, Fairfax, VA

CURRICULUM VITAE

Ning Xu received her Bachelor of Science from Dalian University, Dalian, China in 1993. She was employed as a economist in Bureau of Real Estate in Dalian for five years and received her Master of Science in Systems Engineering from George Mason University in 2004. Ning Xu joined the Center of Air Transportation Systems Research (CATSR) as a research assistant for her Ph.D program in summer 2004. In this capacity, she contributed to research efforts related to Air Transportation Systems. Her research interests include probabilistic modeling and statistical analysis on the system performance of the Air Transportation System.