Estimation and Tactical Allocation of Airport Capacity in the Presence of Uncertainty

by

Varun Ramanujam

Submitted to the Department of Civil and Environmental Engineering
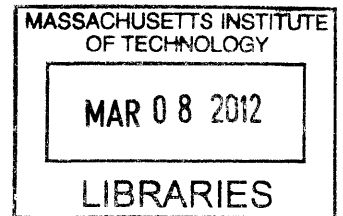in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in the field of Civil and Environmental Systems

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2012

Author . . . .
Department of Civil and Environmental Engineering
Dec 5, 2011

Certified by . . . .
Hamsa Balakrishnan
Assistant Professor of Aeronautics and Astronautics
Thesis Supervisor

Accepted by . . . .
Heidi M. Nepf
Chair, Departmental Committee for Graduate Students

Estimation and Tactical Allocation of Airport Capacity in the Presence of Uncertainty

by

Varun Ramanujam

Abstract

Major airports in the United States and around the world have seen an increase in congestion-related delays over the past few years. Because airport congestion is caused by an imbalance between available capacity and demand, the efficient use of available capacity is critical to mitigating air traffic delays.

A frequently-adopted traffic management initiative, the Ground Delay Program (GDP), is initiated when an airport expects congestion, either because of very high demand or a reduction in its capacity. The GDP is designed to efficiently allocate the limited airport capacity among the scheduled flights. However, contemporary GDP practice allocates delays to arrivals independent of departures, and relies on deterministic capacity forecasts. This thesis designs and evaluates a GDP framework that simultaneously allocates arrival and departure delays, and explicitly accounts for uncertainty in capacity forecasts.

Efficient capacity allocation requires the accurate estimation of available airport capacity. The first module of this thesis focuses on the modeling of airport capacity and its dynamics. A statistical model based on quantile regression is developed to estimate airport capacity envelopes from empirical observations of airport throughput. The proposed approach is demonstrated through a case study of the New York metroplex system that estimates arrival-departure capacity tradeoffs, both at individual airports and between pairs of airports. The airport capacity envelope that is valid at any time depends on the prevailing weather (visibility) and the runway configuration. This thesis proposes a discrete choice framework for modeling the selection of airport runway configurations, given weather and demand forecasts. The model is estimated and validated for Newark (EWR) and LaGuardia (LGA) airports using archived data. The thesis also presents a methodology for quantifying the impact of configuration switches on airport capacity, and applies it to EWR and Dallas Fort Worth (DFW) airports.

The second module of this thesis extends two existing stochastic ground-holding models from literature, the *static* and the *dynamic*, by incorporating departure capacity considerations to existing arrivals-only formulations. These integer stochastic formulations aim to minimize expected system delay costs, assuming uniform unit delay costs for all flights. The benefits of the integrated stochastic framework are demonstrated through representative case

studies featuring real-world GDP data.

During GDPs, the Collaborative Decision-Making framework provides mechanisms, termed *intra-airline substitution* and *compression*, which allow airlines to redistribute slots assigned by ground-holding models to their flights, depending on flight-specific delay costs. The final part of this dissertation considers collaborative decision-making during GDPs in stochastic settings. The analysis reveals an inherent trade-off between the delay costs achieved by the static and the dynamic stochastic models before and after the application of the CDM mechanisms. A hybrid stochastic ground-holding model that combines the desirable properties of the static and dynamic models is then proposed. The performance of the three stochastic ground-holding models under CDM are evaluated through real-world case studies, and the robustness of the final system delay cost reduction achieved by the hybrid model for a range of operating scenarios is demonstrated.

Thesis Supervisor: Hamsa Balakrishnan
Title: Assistant Professor of Aeronautics and Astronautics

Acknowledgments

A doctoral degree is an undoubtedly proud achievement, and marks a culmination of an extended period of personally gratifying endeavor and no less measure of external assistance. In that regard, I owe my foremost share of gratitude to my doctoral committee, comprising of my advisor Prof. Hamsa Balakrishnan and respected Professors Cynthia Barnhart and Amedeo Odoni. They were instrumental in shepherding me through the tricky patches encountered enroute my study. In particular, I am thankful for the levity provided by my advisor, Prof. Balakrishnan, regarding the definition and pursuit of my research agenda. I firmly believe I have benefited a great deal from this experience which has reinforced my mindset and knowhow as I prepare to embark on a research career. My advisor for Master's research, Prof. Moshe Ben-Akiva, was another influential figure in my graduate schooling.

In the hierarchy of benefactors, I am obliged to mention my wondrously supportive labmates next. My three-and-a-half year stint in the ICAT lab has fostered many (un)forgettable moments that now adorn the fondest sections of my memory. A genial environment facilitated by lively and irrepressibly fun-loving peers spanning a gamut of academic, social and cultural backgrounds was the cornerstone of an enriching working experience. While banter and sarcasm were never in meagre supply, it was the occasional gaffes committed in the more solemn moments that I will cherish unreservedly. Here's a heartfelt tribute to fellow veterans Harshad, Ioannis, Nikolas, Claire, Diana, Lishuai, Alex (the "brit" Donaldson) and Hanbong. To Harshad I am profoundly indebted for serving as an untiring sounding board through my quirky musings. He wouldn't be the only person eagerly awaiting my impending deliverance. A special mention is warranted for Sarvee, Olivier, Tomas, Gerasimos and Mehdi for their fleeting but sprightly presence, and to Pierre-Olivier, Lanie and Alexandre Jacquillat for their more recent revival of that freespirited attitude symbolizing the masters program.

My friends outside the lab played a significant role in shaping my social life and helped sustain a healthy environment replete with frivolity that bordered tantalizingly on the wanton, without ever crossing over. Vikrant, Sahil, Murali, Vivek, Raghu, Harshad, Amit, Abhishek are just a few names that come soaring to my mind in this context.

Lastly, and lastingly, I want to thank my parents and my elder brother who have been the staunchest pillars of support through all my undertakings in life. And my undergraduate professor, Dr. K. Srinivasan, who was my first mentor and helped harness my nascent research interests. The values instilled by these people continue to serve as guiding beacons through my darkest hours.

Contents

List of Figures

11

14

List of Tables

Chapter 1

Introduction

The air transportation system is a large, complex, global network that transports people and cargo around the world. It is a key facilitator of global business and tourism. It is estimated that about 10% of travelers worldwide rely upon air transport for at least a part of their trips, and that these services contribute to about 8% of the world's Gross Domestic Product (GDP) through both direct and induced gains [34]. Given its global outreach, the aviation industry impacts a number of stakeholder groups on social, economical, and political fronts. These stakeholders include passengers, air carriers and other businesses associated with passenger and freight traffic, system operators, and environmental groups. The interests of these groups are closely linked to the safe and efficient functioning of the air transportation system.

1.1 Air Traffic Delays

Flight delays are among the most challenging problems faced by the air transportation system, and have been the focus of many studies over the last few decades. A recent NEXTOR report [7] estimated the total cost of delays suffered in the U.S. National Airspace System (NAS) to have exceeded $31 billion in 2007 (Table 1.1). This estimate accounted for both direct and derivative impacts of air transportation delays, with over half of the costs being borne by passengers. While there exist factors like unforeseen mechanical problems that occasionally drive aircraft delays, the most common cause is supply-demand imbalance and

Cost Component	Cost ($ billions)
Costs to Airlines	8.3
Costs to Passengers	16.7
Costs from Lost Demand	2.2
Total Direct Cost	27.2
Impact on Gross Domestic Product	4.0
Total Cost	31.2

Table 1.1: Direct costs of air transportation delays in 2007 [7].

the resulting system congestion. Figure 1-1 presents the breakdown of delay attribution across various causal factors as reported by the Bureau of Transportation Statistics (BTS) [72]. According to this chart, over 90% of the delays are attributed to either a demand surge (Volume) or capacity reduction (Weather). While such imbalances between capacity

■ Weather - 63.45 %

■ Volume - 29.05 %

■ Equipment - 0.01 %

■ Closed Runway - 4.47 %

■ Other - 3.02 %
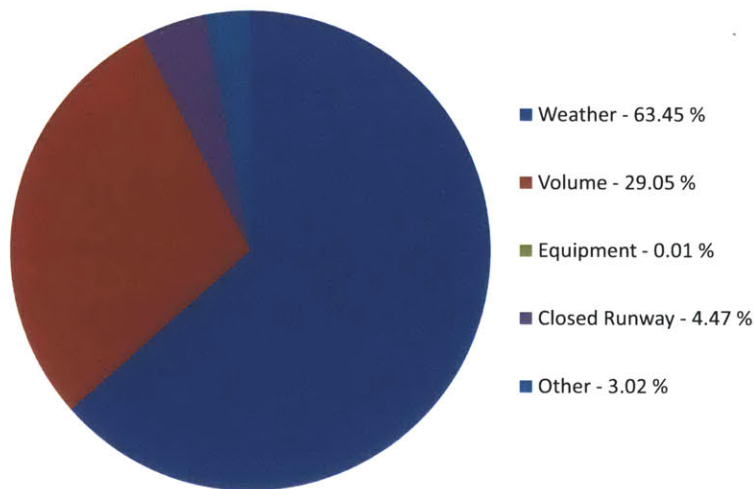
Figure 1-1: Estimated breakdown of air traffic delays by causal factors (courtesy Bureau of Transportation Statistics [72]).

and demand are localized in terms of temporal and geographical scope, they can propagate delays across the air traffic system due to the underlying network connectivity and generate system-wide impacts, as discussed by Pyrgiotis (2011) [64]. With a growth in demand and escalating costs of delays due to system overloading, the effective utilization of available

airport and airspace capacity has become more important in recent years. In the NAS, the supply-side element that is usually constrained is the airport capacity. This is in contrast to Europe, where enroute airspace serves as a typical bottleneck, as noted by Lulli and Odoni (2007) [52].

1.2   Delay Mitigation Techniques

Given that the major causes of air delays are congestion-related, potential solutions can be grouped into two broad categories: infrastructure enhancements, and operational refinements. The first class of measures involves increasing the system capacity through the development of new airports, runways or airspace routes. However, in addition to being very expensive, most of these approaches and their associated benefits materialize over a long time-frame. The gains from such measures are also likely to be offset by a rise in NAS demand stimulated by system capacity expansion [21].

By contrast, operational refinements try to improve system functioning so that existing capacity is used more effectively. This class of solutions includes schemes for regulating capacity and demand, that proactively or reactively mitigate delays and complement the existing Air Traffic Management (ATM) infrastructure. This family of congestion management approaches can be further classified as Demand Management or Air Traffic Flow Management (ATFM) measures [21].

Demand management initiatives are strategic economic and administrative policies that seek to alter the magnitude or temporal characteristics of air traffic demand. Proposed approaches include slot restrictions at major airports, congestion pricing, and slot auctions for airport capacity allocation. ATFM measures are tactical schemes that are invoked on a day-to-day basis, as needed, to tackle short-term air traffic congestion by effective redeployment of air traffic flows. They represent a popular area of research because of the relatively low cost of implementation, and have generated a large body of literature over the recent decades. A number of such initiatives are currently practiced to varying degrees of sophistication around the world. Prominent examples include Ground Delay Programs (GDPs), Airspace Flow Programs (AFP), aircraft rerouting, etc.

This thesis focuses on ATFM methods related to airport capacity utilization, since airport capacity is believed to be the most frequent bottleneck in NAS operations [52]. The broad theme of this dissertation is the development of novel approaches to airport capacity management. The next section presents an overview of the state-of-art ATFM designs for airport capacity management in practice and in research literature.

1.2.1 Ground Delay Programs (GDPs)

A frequently-adopted ATFM practice for managing airport capacity is the Ground Delay Program (GDP). GDPs are invoked when an airport expects congestion, either because of very high demand or a reduction in its arrival capacity (due to adverse weather or maintenance-related closures), and are used to ration the limited capacity by assigning ground-holds to the scheduled arrivals in an efficient and equitable way. The underlying principle of a GDP is that by delaying a flight's departure at its origin airport, it can avoid airborne delays on arrival at its overloaded destination airport. This approach is reasonable because ground delays are safer and less expensive than airborne delays.

In the NAS, GDPs are one of the most frequently adopted ATFM measures. During the year 2006, more than 1200 initiations were recorded across all airports, accounting for total flight delays in excess of 13 million min [24].

1.2.1.1 GDP example

Consider an airport with good weather arrival capacity of 30 operations per hour (2 min inter-operation separation), that is expected to be halved in the immediate future. Given a scheduled demand comprising of uniformly spaced arrival slots from 11:00 am to 11:10 am as depicted in Table 1.2, the delayed arrival slots and the corresponding ground delays assigned by a GDP to conform to the reduced capacity are as shown in Table 1.2.

Note that the scheduled arrival order is preserved following ground delay allocation. This is in compliance to the Ration-by-Schedule (RBS) principle observed during GDPs, which ensures fairness in the allocation of airport capacity. Also, flights that are already airborne at the time of GDP initiation are exempt from delay allocation, in order to eliminate avoidable

Airline-Flight	Scheduled Arrival Time	GDP Arrival Slot	Assigned Ground Delay (min)
A1	11:00	11:00	0
B1	11:02	11:04	2
C1	11:04	11:08	4
A2	11:06	11:12	6
B2	11:08	11:16	8
A3	11:10	11:20	10

Table 1.2: Illustrative GDP example (5 flights from 1100-1110 hours).

airborne delays.

1.2.2   Collaborative Decision-Making (CDM)

Another aspect of the current GDP framework is the philosophy of collaborative decision-making (CDM). This paradigm seeks to encourage airline participation in the delay allocation process, by providing two mechanisms through which airlines can respond to the GDP slot allocation determined by the airport. The first mechanism, called intra-airline slot substitution, allows an airline to swap GDP-assigned slots between its flights, and thereby rearrange its schedule in accordance with flight-specific delay costs [18]. Revisiting the GDP example from Table 1.2, and focusing on airline A's flights alone, consider a situation in which flight A3 is considered more delay-sensitive than flight A2. Through the CDM mechanism of intra-airline substitution, airline A can swap assigned slots for flights A2 and A3 after the original GDP allocation, as shown in Table 1.3. Such information on flight-specific delay cost need not be revealed to the system operator, which treats all flights equally in the original RBS-driven allocation. Intra-airline substitution therefore helps airline further reduce internal schedule costs.

The second mechanism managed under CDM is that of Compression [18]. Through this mechanism, an airline can voluntarily release an assigned GDP slot if no flight amongst its fleet can feasibly occupy it. The system handles the newly-vacated slot by advancing flights from later slots to fill up the gap, thereby "compressing" the schedule, and allots the first feasible slot to the airline that released the slot. In this manner, this mechanism incentivizes truthful airline participation in the compression process, which in turn helps restore efficiency

(a) Original GDP allocation

Airline-Flight	Scheduled Arrival Time	GDP Arrival Slot	Ground Delay (min)
A1	11:00	11:00	0
B1	11:02	11:04	2
C1	11:04	11:08	4
A2	11:06	11:12	6
B2	11:08	11:16	8
A3	11:10	11:20	10

(b) After airline A's slot swap

Airline-Flight	Scheduled Arrival Time	GDP Arrival Slot	Ground Delay (min)
A1	11:00	11:00	0
B1	11:02	11:04	2
C1	11:04	11:08	4
A3	**11:10**	**11:12**	**2**
B2	11:08	11:16	8
A2	**11:06**	**11:20**	**14**

Table 1.3: Intra-airline slot substitution example.

in slot utilization. In the GDP example from Table 1.2, consider a situation in which flight B1 is now expected to suffer 10 mins of mechanical delay beyond its original schedule, and can only arrive at the earliest by 11:12 a.m. instead of 11:02 a.m, and hence cannot feasibly use its assigned arrival slot at 11:04 a.m. Since B2 cannot be feasibly advanced to occupy this slot either (since its earliest arrival time is 11:08 a.m.), airline B forfeits the 11:04 a.m. slot. Through Compression, the subsequent flights C1 and A2 are moved up in sequence to occupy the vacated slot, and to open up the slot at 11:12 a.m. for B1, as shown in Table 1.4.

Recently, the Compression process has been increasingly replaced by its upgraded equivalent, the Slot Credit Substitution (SCS) [3]. SCS is a real-time, dynamic version of Compression that is executed in an event-driven manner in response to the most recent slot forfeiture, while the earlier version was a batch process that was executed periodically, and handled multiple forfeitures simultaneously.

The flow-chart in Figure 6-1 illustrates the GDP/CDM operational framework, comprising of three main modules executed in sequence. The first module pertains to the development of capacity forecasts for the short-term horizon over which the GDP is declared. This

(a) Original GDP allocation

Airline-Flight	Scheduled Arrival Time	GDP Arrival Slot	Ground Delay (min)
A1	11:00	11:00	0
B1	11:02	11:04	2
C1	11:04	11:08	4
A2	11:06	11:12	6
B2	11:08	11:16	8
A3	11:10	11:20	10

(b) After Compression

Airline-Flight	Scheduled Arrival Time	GDP Arrival Slot	Ground Delay (min)
A1	11:00	11:00	0
C1	**11:04**	**11:04**	**0**
A2	**11:06**	**11:08**	**2**
B1	**11:12**	**11:12**	**0**
B2	11:08	11:16	8
A3	11:10	11:20	10

Table 1.4: Compression example.

is followed by the second (GDP) module which determines ground-holds and slot allocations based on the capacity forecasts, and the third module, which contains the CDM mechanisms implemented in response to the slot allocation.
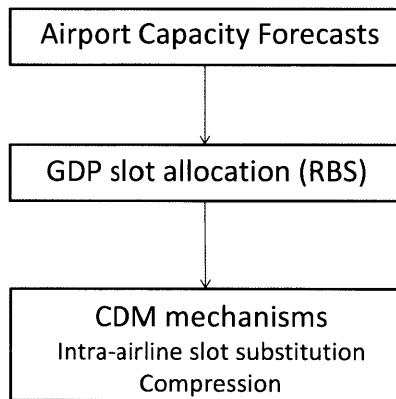
Figure 1-2: Framework for GDPs in the CDM paradigm.

The introduction of CDM mechanisms is considered a resounding success, and has resulted in significant delay cost reductions for airlines [18]. Given its prominence and widespread

acceptance as an effective congestion mitigation measure, the GDP framework has prompted an extensive body of literature that seeks to address any drawbacks, and to propose further refinements. The next section highlights the key limitations of the current GDP framework that motivate the research presented in this thesis.

1.2.3 Limitations of current GDP designs

1.2.3.1 Arrival-centric approach

The current design for GDP caters only to arrival operations at the congested airport, as shown in Figure 1-2. The prediction of capacity, its allocation as well as CDM-based revision of this allocation is performed on the arrival elements alone. Departures are not systematically managed by a GDP, but are instead accommodated on an ad hoc basis. In other words, since simultaneous departure and arrival capacities at an airport are interdependent [21], scheduled take-offs from the congested airport during a GDP are serviced in each time interval contingent on allotted arrival slots and available airport capacity. Such an explicit prioritization of arrivals stems from the goal of avoiding arrival airborne delays at the expense of departure ground delays at the congested airport. However, this prioritization might not be beneficial for the entire system during periods with comparable arrival and departure demands. During GDPs recorded in the year 2006, there was an average of 0.7 departures scheduled per arrival at the affected airports [24]. This observation suggests that there is significant scope for realizing systemic delay benefits through the integrated management of arrival and departure operations during GDPs.

1.2.3.2 Deterministic approach

Current GDP processes of slot allocation and CDM mechanisms rely on deterministic capacity forecasts. This feature is not supported by empirical evidence, in which the actual airport capacity profile experienced during a GDP is seldom predicted accurately at its onset. In practice, the entire sequence of GDP processes are revisited with every update in capacity forecasts. Each new iteration is constrained by irreversible decisions regarding slot utilization made by airports and airlines based on earlier forecasts, meaning there is a loss

26

in delay mitigation efficiency with every GDP revision. Each initiated GDP in 2006 was revised 2.5 times on average before its termination [23], implying that capacity uncertainty is a serious problem.

Models for jointly allocating arrival and departure capacity have been studied for single-airport [31, 32] and multi-airport [74, 13] settings; these approaches have been limited to a deterministic capacity framework. Studies in literature have considered stochastic capacity forecasts comprising of multiple capacity scenarios and associated probabilities at the outset of a GDP. This results in a stochastic programming formulation that derives a ground-hold allocation that explicitly plans for uncertainty in future capacity. Depending on the nature of the resulting ground-hold allocation, there are static stochastic models, proposed first by Richetta and Odoni (1993) [67], that determine a fixed solution across all capacity scenarios, and dynamic stochastic models, proposed by Richetta and Odoni (1994) [68] and enhanced by Mukherjee and Hansen (2007) [55], that allow for ground-hold revisions contingent on scenario materialization.

These stochastic ground holding models, however, retain the arrival-centric focus of GDP design, and exclude departures from the ground-holding framework. Another aspect that has not been previously researched in detail is the design of CDM mechanisms under uncertainty, that is, the extension of current deterministic frameworks for intra-airline slot substitution and compression to deal with stochastic ground-holding solutions. To the best of our knowledge, Mukherjee (2007) [57] is the only study that has considered this question.

1.3   Thesis Objectives

The thesis aims to integrate arrival and departure considerations, and explicitly handle uncertainty within the existing GDP framework, and to study potential system delay benefits attainable through these enhancements. In particular, the dissertation addresses the following challenges related to the GDP processes shown in Figure 1-2:

1. Characterization of airport capacity, including the estimation of airport capacity dynamics.

27

2. Allocation of airport capacity in the presence of uncertainty.

3. Development of CDM mechanisms that can accommodate uncertainty.

1.3.1 Airport capacity characterization

Given that the first step in a GDP is to generate reliable capacity forecasts, this research looks to characterize key aspects pertaining to an airport's arrival and departure capacities, and thereby construct a comprehensive model for airport capacity. The focus is on developing statistical models that can utilize empirical data on airport operations to estimate airport capacity tradeoffs, and to predict the evolution of capacity in the near future.

1.3.2 Airport capacity allocation under uncertainty

This research considers the integrated management of arrival and departure operations within a stochastic capacity framework for a GDP. This represents a simultaneous rectification of the two previously mentioned drawbacks in the current GDP process of capacity allocation, which have previously only been addressed separately. The significant volumes of scheduled departures relative to scheduled arrivals during recorded GDPs, along with the unexplored topic of airport capacity sharing under uncertainty, provide motivation for the investigation of system delay benefits from integrated capacity management.

1.3.3 CDM mechanisms under uncertainty

The final objective of this thesis is to extend CDM mechanism designs to a stochastic GDP framework. Given the benefits of stochastic models over deterministic designs for determining ground-holds, the design of CDM mechanisms compatible with stochastic ground-holding solutions is important. Any new proposed models as well as existing stochastic ground-holding models [67, 55] need to be studied and compared in conjunction with CDM mechanisms, in terms of the ultimate system delay benefits.

1.4 Thesis Contributions

This dissertation develops modeling frameworks to address the objectives described above, and applies them to representative case studies.

1.4.1 Airport capacity modeling

1.4.1.1 Quantile regression models for capacity envelope estimation

This thesis proposes a quantile regression-based approach to estimate airport capacity envelopes using observed throughputs and other operational factors at an airport. The advantages of the proposed approach over existing techniques is demonstrated. Along with tradeoffs between arrival and departure capacities at an airport, this statistical model can also estimate interdependencies between operational capacities at neighboring airports. The latter relationship is a widely acknowledged phenomenon in metroplexes, that is, a cluster of closely located airports that interact through shared terminal airspace. The proposed methodology has been applied to the three main airports in the NY metroplex: EWR, JFK and LGA, and the influence of underlying factors like runway configurations, visibility, etc. on capacity envelopes are statistically quantified.

1.4.1.2 Discrete choice models for runway configuration selection

The key factors driving the capacity tradeoffs at an airport are the visibility and the choice of runway configuration. While the visibility is related to weather conditions, the processes that determine the choice of runway configuration have received little attention. In this thesis, empirical models are developed to study runway configuration dynamics. The process of configuration selection is modeled using a discrete choice framework. The estimated model is applied to EWR and LGA airports to uncover the effect of several factors, including weather, demand, etc. on configuration selection.

1.4.1.3 Two-stage regression for configuration switch effects

This thesis also features the first effort to systematically estimate the impact of configuration changes on the operational capacity of an airport, a phenomenon that has been assumed in a number of past studies. A two-stage linear regression framework is designed for this task, and applied to EWR and DFW airports to extract relevant details on configuration switch effects.

The combination of models for estimating capacity envelopes and configuration selection decisions constitute a comprehensive framework for modeling the dynamics of airport capacity.

1.4.2 Integrated airport capacity allocation under uncertainty

Models for simultaneously allocating arrival and departure capacity under stochastic capacity forecasts are developed by incorporating departure considerations into arrival-centric stochastic ground-holding models. Integrated versions are also developed for existing "Static" and "Dynamic" models, and applied to representative case studies featuring hypothetical as well as real-world data of scheduled demand and airport capacity profiles during GDP. Delay benefits over the existing arrival-centric approaches are demonstrated and analyzed. The results indicate that typical real-world GDPs stand to gain substantially by adopting an integrated approach to capacity allocation.

1.4.3 CDM mechanisms under uncertainty

1.4.3.1 Stochastic designs for CDM mechanisms

Extensions are designed for the CDM mechanisms of "Intra-airline slot substitution" and "Compression" in the form of integer formulations that render them compatible with stochastic slot allocations, while retaining the original principles of their deterministic counterparts. A detailed overview of the slot-specific information exchanged between the airports and airlines to facilitate the execution of CDM mechanisms in a stochastic framework is also provided.

1.4.3.2 A new stochastic ground-holding model suited to CDM

The amenability of the solutions from the "Static" and the "Dynamic" stochastic models to CDM mechanisms is explored, and an interesting trade-off between the dynamic models' ground-hold efficiency and the static models' flexibility for slot substitution is established. This trade-off critically influences the relative performances of the two stochastic models in a complete GDP framework combining slot allocation and CDM mechanisms. A "Hybrid" stochastic model that combines the favorable properties of both models is developed, and its ability to achieve an effective compromise in the final system delay costs is demonstrated using both hypothetical and real-world case studies. The results underline the value of the hybrid model as a robust option for ground-hold allocation within a stochastic GDP framework. The properties of the stochastic integer formulation for the hybrid stochastic model are also analyzed, and some significant results concerning its computational tractability are derived.

1.5 Organization of the Thesis

The organization of this dissertation is as follows. Chapter 2 discusses the problem of airport capacity estimation and the quantile regression-based methodology proposed for the same. The material presented in this chapter previously appeared in Ramanujam and Balakrishnan (2009) [65]. Chapter 3 presents the statistical models for configuration selection and switch effect estimation, and was contains material that was presented in Ramanujam and Balakrishnan (2011) [66]. Having completed the study of capacity estimation, Chapter 4 proceeds to analyze the problem of allocating integrated airport capacity under uncertainty and the associated delay benefits through relevant case studies. Chapter 5 examines the amenability of the static and dynamic stochastic ground-holding models to CDM mechanisms, and describes the formulation of the hybrid stochastic model and the related results on its computational tractability. Chapter 6 continues the discussion by developing the stochastic designs for the CDM mechanisms and evaluating the performances of the three ground-holding models in a stochastic GDP framework through appropriate case studies. The final chapter summarizes the thesis contributions and outlines promising directions for

future research.

Chapter 2

Estimation of Airport Capacity Envelopes

2.1 Introduction

As noted in Chapter 1, airports and terminal airspace regions are usually the capacity-limiting components during periods of air traffic congestion in the US. The accurate estimation of airport capacity is therefore vital not only for efficient planning of landing and takeoff operations, but also for mitigation of congestion-induced delays. This is particularly significant given the predicted increase in demand in the NextGen system. The effects of congestion are beginning to be seen even in today's system: in the United States, between 2006 and 2007, there was a 30% rise in delays due to terminal-area volume, while there was only a 1% increase in traffic [17, 24]. A critical means of tackling the expected growth in demand is through improved utilization of airport capacity, especially in congested terminal-areas. With the emergence of secondary and even tertiary airports in the most congested regions of the country (for example, the New York area, the San Francisco Bay area, the Los Angeles area, Boston and Chicago [16]), the problem of coordinating operations in multi-airport systems to use terminal-area capacity more efficiently has also become increasingly important.

The operating capacity of an airport is given by its arrival capacity (the number of aircraft landings per hour) and its departure capacity (the number of departures per hour). The

arrival or departure capacity is primarily limited due to the temporal separation requirements imposed by FAA guidelines between successive operations. The inter-operational separations are meant to avert potential physical conflicts when using airport resources. Due to the shared nature of ground resources such as runways and taxiways, there is a tradeoff between the simultaneous arrival and departure capacity at an airport. This phenomenon arises due to the need to observe the above-mentioned separations between successive pairs of arrivals and departures that operate on interacting runways. The conceptual basis of this tradeoff has been widely acknowledged in literature, including studies of Blumstein (1959) [15] and Janic and Tosic (1982) [43].

The tradeoff between an airport's arrival and departure capacity is quantified using the concept of capacity envelope. An airport capacity envelope is the boundary (generally approximated as a convex polygon on the plane with the arrival and departure rates as axes) that defines the envelope of the maximum capacities that can be achieved under specified operating conditions, and captures the tradeoff between the maximum arrival and departure rates [21]. The operating conditions influencing the tradeoff encompass factors such as the relative alignment of arrival and departure runways (defined as the runway configuration), meteorological factors like wind and visibility, the aircraft fleet mix, etc. Each of these factors dictate the required inter-operational separations that in turn determines the operational capacities. Figure 2-1 illustrates the representative capacity envelopes for an airport for an arbitrary runway configuration split between two visibility categories typically defined for operations.

In the case of an isolated airport, the arrival departure capacity tradeoffs would be sufficient to describe its operational characteristics under any given set of exogenous conditions. However, this is not the only form of operational tradeoff that defines airport capacity in many contemporary settings. The emergence of several clusters of core, secondary and regional airports (known collectively as a *metroplex* [44]) results in the complex interaction of traffic flows within an already congested airspace, as shown in Figure 2-2. These interactions between arrival and departure operations at proximate airports could potentially make it infeasible to simultaneously operate all the airports at their individual optimal runway configurations or the corresponding operational capacities. While simultaneous arrivals and

34
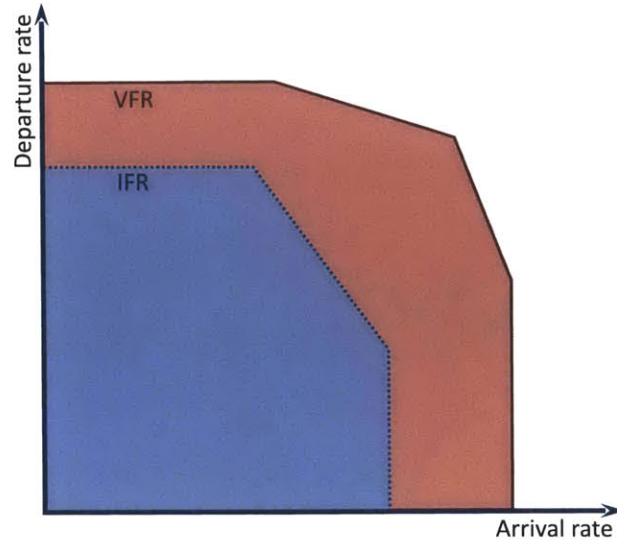
Figure 2-1: Illustration of capacity envelope for an airport, under a particular runway configuration, for different meteorological conditions: (1) Visual Flight Rules (VFR), and (2) Instrument Flight Rules (IFR). The shaded regions represent the feasible operating points.

departures at a given airport conflict due to shared ground resources, simultaneous operations from neighboring airports conflict due to shared airspace. This implies that the process of allocating the capacity of an airport at a given time would have to not only be mindful of the arrival-departure tradeoffs for that airport, but also the tradeoffs with arrival and departure capacities at the other airports within the multi-airport system. Given the nature of the tradeoffs, the concept of capacity envelope can be extended to represent the inter-relationships between simultaneous operational capacities at neighboring airports. While there is anecdotal and descriptive evidence of such interactions between air traffic flows into and out of neighboring airports, there have been no attempts to quantify these interactions and their impact on capacity using operational data.

A detailed understanding of airport capacity tradeoffs and their dependence on external factors such as operating conditions and airport layout, both in the single- and multi-airport settings, is necessary for the efficient utilization of airport capacity. Within the context of this thesis, such an understanding is a prerequisite for optimal capacity allocation during GDPs. While operational tradeoffs have not been studied so far in the multi-airport setting, the estimation of single-airport capacity envelopes has traditionally been restricted to theoretical approaches that rely upon an abstracted model of airport operations as decribed in [15, 37,
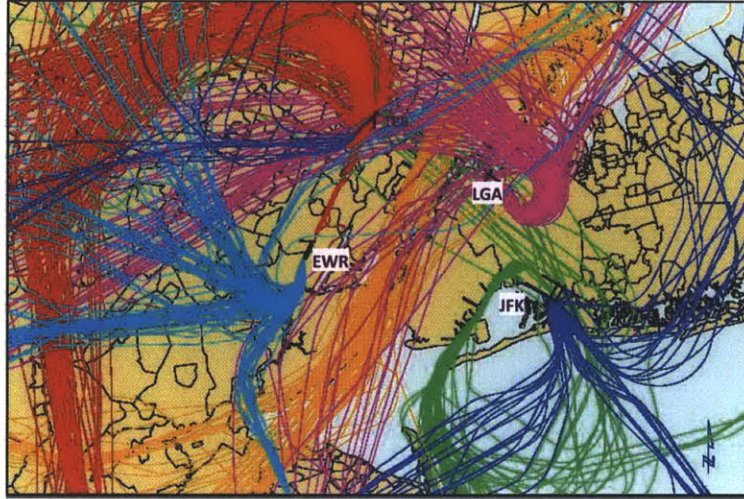
Figure 2-2: The New York airspace, showing the intertwined arrival and departure routes into the core airports. Figure courtesy of the Port Authority of New York and New Jersey (PANYNJ).

40]. Empirical methods that use operational data have only been employed in recent times, but have generally been ad hoc approaches like the one proposed by Gilbo (1993) [31]. This approach serves as the precursor for that proposed in this research, and is discussed in the following sections of this chapter.

This chapter develops a systematic statistical technique, based on quantile regression, to estimate piecewise-linear, convex capacity envelopes using observed throughput data, for both individual airports as well as the corresponding inter-airport dependencies in a multi-airport system. The problem of estimating the capacity envelopes as quantiles (or percentiles) of the empirical data is formulated as a linear program. The proposed technique is illustrated using reported data on arrivals and departures at the three major New York area airports, namely John F. Kennedy (JFK), Newark (EWR) and LaGuardia (LGA), for the years 2006 and 2007. The analysis identifies the key factors that influence the intra- and inter-airport capacity tradeoffs, and determines the associated capacity envelopes.

The output of the capacity estimation process studied here serves as a key input for the models of capacity allocation discussed in later chapters.

2.2 Related Literature

Airport capacity is affected by various external factors such as air traffic controller procedures and pilot behavior, approach and departure speeds, runway and taxiway occupancy times, weather, etc. Theoretical approaches to capacity estimation have traditionally modeled these factors through simplified models of aircraft behavior, and derived the capacity using the mandated separation time between successive aircraft operations [15, 37, 40, 71, 50, 73]. These theoretical models construct the capacity envelope through linear interpolation between capacity values computed at specific arrival/departure mix ratios. Newell (1979) [58] and Odoni et al. (1997) [60] provide comprehensive reviews of contemporary analytical and simulation methods that adopt the above approach.

Empirical estimation approaches have the potential to reflect the practical operating thresholds achieved by controllers at airports, in contrast to idealized models of capacity tradeoffs. Recognizing this, Gilbo (1993) [31] proposed a quasi-statistical procedure for estimating the capacity envelope of a single airport under specific runway configurations. 15-min arrival and departure counts were used to estimate the capacity envelope as the convex hull (polygon) of the planar scatter of the counts. Frequency-based filtering was employed on the outer perimeter of the data to eliminate outliers, that is, data points that were deemed to be an unreliable representation of airport capacity. Alternative outlier rejection criteria such as proximity to neighboring observations and rank order statistics were also mentioned. Figure 2-3 schematically illustrates the functioning of Gilbo's (1993) method for estimating capacity envelopes.

In contrast, this paper proposes a systematic statistical approach based on the principle of *quantile regression* for estimation of intra- and inter-airport capacity envelopes from observed data. On the statistical estimation front, quantile regression attempts to determine statistics such as the median or a general percentile of the dependent variable as a function of independent variables from a given sample of observations [46]. This differs from the regular least-squares regression which estimates the mean of the dependent variable instead. Koenker (2001) provides an elaborate description of the mathematical framework underlying quantile regression, while also highlighting some significant applications from literature.
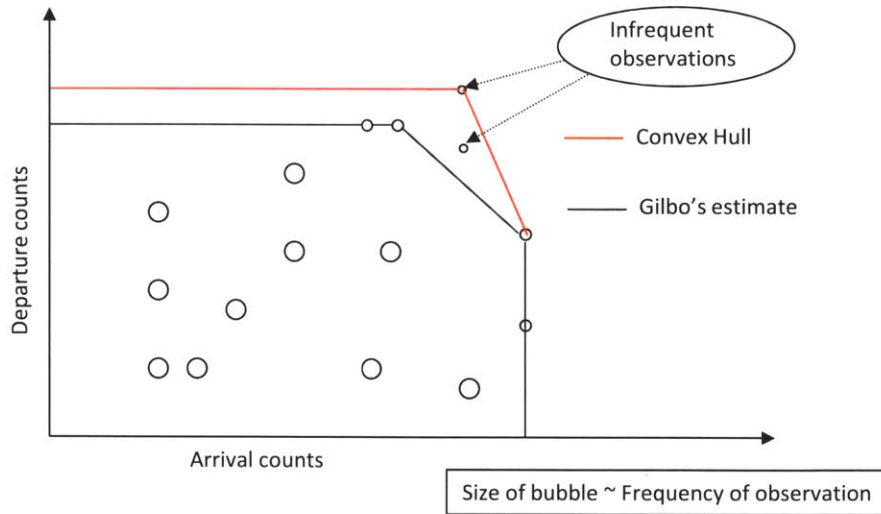
Figure 2-3: Gilbo's methodology for capacity envelope estimation

Amongst these, there exist some precedents that attempt to model capacity functions as quantiles. For example, Bernini et al. (2004) modeled the production frontier in classical economics as a higher-order quantile (90−100%ile), and studied the interaction between underlying determinants at intermediate quantiles [11].

This research extends the quantile regression technique to the case of airport capacity envelope estimation. The capacity envelope, which represents the inter-relationship between arrival and departure capacity, is examined as a departure capacity function conditioned on arrival counts. This capacity function is estimated as a higher-order quantile, with the chosen order of quantile serving as a replacement for the frequency filter adopted in Gilbo (1993) [31]. The formulated approach looks to estimate piecewise-linear, concave quantile functions of departure capacity to generate the capacity envelope. The use of quantile regression enables, through hypothesis testing, explicit study of the influence of exogenous attributes on the capacity curve. These attributes may include geometric details of runway configurations, weather conditions like visibility and wind, etc. In this manner, the quantile regression approach helps obtain a comprehensive estimate for airport capacity.

The next sections of this chapter outline the methodological framework of the quantile

regression approach and its application to estimate the intra- and inter-airport capacity tradeoffs in the NY multi-airport system.

2.3   Problem Statement

Given the counts of arrival and departure throughputs at an airport per time interval (say, 15 minutes), the task is to determine the capacity envelope enclosing the set of observations by suitably excluding outliers. A related goal is to identify key factors that affect the shape of the capacity envelope and estimate their influence.

The proposed approach to solving this problem draws from the field of quantile regression. A data point is said to be at the τth quantile if it is larger than a proportion $\tau/100$ of the data points, and less than a proportion $(1 - \tau/100)$ [46]. Similar to least-square regression techniques that estimate the mean of a response variable given values of the predictor variables, quantile regression techniques estimate other statistics, such as the median or a quantile [46]. Since the airport capacity envelopes represent the upper limits of operating capacity, quantile regression techniques, with τ sufficiently large, are suitable mechanisms for estimating them. In other words, by setting $\tau = 99.5$, one could look to construct the capacity envelope such that 99.5% of all reported operating points fall within the feasible region. The remaining 0.5% would constitute the list of outliers. Thus, the order of quantile (value of τ) chosen to represent the airport capacity is conceptually similar to the frequency-based filter adopted by Gilbo (1993) [31] to discard spurious data (for example, reporting errors).

Unlike least squares regression, the estimation of parameters of a quantile regression function is conducted by solving a linear program which attempts to minimize sum of asymmetrically weighted absolute deviations [46, 47]. While quantile regression has traditionally been used to determine linear quantile functions, in the case of airport capacity envelopes the focus is on determining a piecewise-linear, concave, continuous function that represents the quantile. The concavity assumption is in accordance to the accepted notion in literature that capacity envelopes are always convex, that is, increasing the number of arrivals impacts the departure capacity with monotonically steeper magnitude. Appendix II describes how this

convexity assumption can be tested for its statistical validity over a given set of observations using the quantile regression-based estimation framework.

In this chapter, convexity of capacity envelope is assumed. The standard linear programming formulation for quantile regression estimation is extended to include constraints enforcing continuity and concavity across the linear pieces. In addition to airport-specific capacity envelopes, the proposed approach is used to study the presence and magnitude of tradeoffs between operations at neighboring airports, and the results are documented in a later section of this chapter.

2.4 Modeling Framework

This section describes the modeling of capacity envelopes, and the formulation of the linear programs to estimate them.

2.4.1 Capacity envelope representation

The capacity envelope representation is decomposed into two parts as depicted in Figure 2-4: an arrival rate threshold (highlighted in green) and a departure capacity function (highlighted in red).
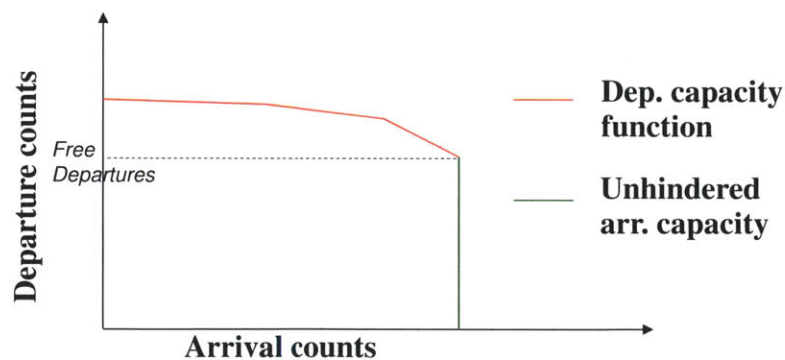
Figure 2-4: Decomposed representation of capacity envelope.

For the departure capacity function, the arrival rate is treated as the independent variable, and the departure capacity for any time interval is estimated as a function of the arrival

40

rate using quantile regression. This decision to model departure capacity as the dependent variable is based on the observation that arrivals are generally given priority at airports over departures. It is therefore reasonable to assume that while the arrival rate depends on the demand, departure capacity is traded off to accommodate arrivals. In other words,

$$\text{Departure capacity} = f(\text{arrival rate}). \qquad (2.1)$$

As shown in the figure 2-4, the arrival rate threshold is defined as the maximum number of arrivals that can be accommodated in a time period for a given set of runway configuration and weather conditions. It is also called the *unhindered arrival capacity* since it spans operating points at which the presence of departures do not reduce the arrival rate, i.e., the number of departures are within the "free departure" rate. It corresponds to the *arrivals-only point* on the capacity envelope [21], and is estimated independently of the departure capacity.

Because arrival and departure counts are reported for 15-minute intervals, the counts are used as measures of the arrival and departure rates. Since the capacity envelope is assumed to form a convex polygon, the departure capacity needs to be estimated as a piecewise-linear, concave quantile function of the arrival rate for the range of observed arrival rates. Recent research on piecewise quantile regression models, like Kim (2007) [45] and Furno (2007) [28], have assumed no knowledge of the potential locations of the breakpoints where the slope of the quantile function changes. In this study, one can leverage upon knowledge of the structure of the capacity envelope to estimate the piecewise linear function. There is a finite number of potential breakpoints on the capacity envelope due to the fact that the arrival count (in a 15-minute interval) is a nonnegative integer, and is bounded to be within a manageable threshold (typically less than 20). This allows the estimation of the linear segments of departure capacity over all unit intervals of arrival count with minimal computational effort. The piecewise-linear representation for the departure capacity function then has the following form:

$$Q_\tau^{\text{dep}}(y|x) = \sum_i \alpha_k^i \theta^i + \left(\sum_i \beta_k^i \theta^i\right)x, \text{ for } (k-1) \le x \le k, \ \forall k \qquad (2.2)$$

41

where

y is the departure count,

$x \in \{0, 1, \cdots, x_{\max}\}$ is the arrival count,

$Q_\tau^{\mathrm{dep}}(y|x)$ is the τ-quantile function of y with respect to x, which represents the departure capacity function,

$k \in \{1, ... x_{max}\}$ denotes the kth interval of the arrival count,

$\theta^i \in \theta$ are the factors influencing the departure capacity function (for example, VFR/IFR conditions or runway alignment), and

α_k^i and β_k^i are the intercept and slope contributions of the factor θ^i, for the kth linear segment.

The unhindered arrival capacity is a scalar value, and is estimated as a high-order quantile of the observed arrival counts x in the dataset. This quantile measure can be represented as a function of influencing factors similar to the above expression for the departure capacity function.

$$Q_\tau^{\mathrm{arr}}(x) = \sum_i \gamma^i \phi^i \tag{2.3}$$

Here, γ^i serves as the coefficient of the factor ϕ^i, where the set of factors ϕ is the counterpart of θ in the quantile specification for the departure capacity function. As mentioned earlier, the estimation for the unhindered arrival capacity is independent of that for the departure capacity function. The formulations for these estimation tasks are described below.

2.4.2   Formulation of estimation problem

Given a data set of N operating observations involving y, x and θ at a given airport, the process of estimating the piecewise coefficients (α_k, β_k) for a chosen quantile τ for the departure capacity function Q_τ^{dep} involves solving the following linear optimization problem:

$$\text{Minimize} \quad \sum_{n=1}^{N} Z_n \qquad\qquad (2.4)$$

subject to:

$$Z_n \geq y_n - \left[\sum_i \alpha_k^i \theta_n^i + (\sum_i \beta_k^i \theta_n^i)x_n\right] \qquad \text{if } k-1 \leq x_n \leq k, \forall n \qquad (2.5)$$

$$Z_n \geq \omega_\tau \left[\sum_i \alpha_k^i \theta_n^i + (\sum_i \beta_k^i \theta_n^i)x_n - y_n\right] \qquad \text{if } k-1 \leq x_n \leq k, \forall n \qquad (2.6)$$

$$\sum_i \beta_{k+1}^i F_i(\theta_n^i) \leq \sum_i \beta_k^i F_i(\theta_n^i), \ \forall k \in \{1, \cdots, x_{\max} - 1\},$$

$$\forall F_i(.) \in \{\max_n(\cdot), \min_n(\cdot)\}, \ \forall i \qquad (2.7)$$

$$\sum_i \beta_1^i F_i(\theta_n^i) \leq 0, \ \forall F_i(\cdot) \in \{\max_n(\cdot), \min_n(\cdot)\}, \ \forall i \qquad (2.8)$$

$$\sum_i \alpha_k^i F_i(\theta_n^i) + [\sum_i \beta_k^i F_i(\theta_n^i)]k = \sum_i \alpha_{k+1}^i F_i(\theta_n^i) + [\sum_i \beta_{k+1}^i F_i(\theta_n^i)]k,$$

$$\forall k \in \{1, 2, ..., x_{\max} - 1\}, \ \forall F_i(.) \in \{\max_n(.), \min_n(.)\}, \ \forall i \qquad (2.9)$$

In the above formulation, constraints (B.2)-(B.3) define the absolute deviation of the estimated quantile function for departure capacity from the observed value for departure count y. Constraints (2.7)-(2.8) help ensure concavity and non-positivity of the piecewise slope estimates for all values of $\theta^i \in \theta$ in the range $[\theta^{i,\min}, \theta^{i,\max}]$. As mentioned earlier, these constraints are motivated by the accepted notion that the magnitude of arrival-departure tradeoff monotonically increases with arrival rate. Constraints (B.5) ensure continuity of adjacent segments. $\omega_\tau = (100 - \tau)/\tau$ is the asymmetric weight applied upon the negative deviations (constraint (B.3)), where τ denotes the order of the quantile estimated. This feature of asymmetric weighting of deviations is the central estimation principle of quantile regression [47].

The size of the above formulation depends on the number of observations in the data set (N), the maximum arrival count in a 15-minute interval ($x_{\max}$) and the number of factors considered in the representation of departure capacity function (that is, the size of vector θ).

The unhindered arrival (or arrivals-only) capacity can be estimated using a simplified version of the above formulation. The data set is restricted to observations where the arrival

43

rate is not impeded by departures. That is, for observations where the departure count is within the free departure rate obtained from the departure capacity estimation above. The free departure rate, as depicted in Figure 2-4, is the lowest departure capacity across the range of arrival counts. The unhindered arrival capacity is estimated by solving the following linear program:

$$\text{Minimize} \quad \sum_{n=1}^{N} Z_n \tag{2.10}$$

$$\text{subject to:}$$

$$y_{n,\min}^{\text{cap}} = \sum_i \alpha_{x_{\max}}^i \theta_n^i + \left(\sum_i \beta_{x_{\max}}^i \theta_n^i\right) x_{\max} \tag{2.11}$$

$$\left. \begin{array}{l} Z_n \geq x_n - \sum_i \gamma^i \phi_n^i \\[2em] Z_n \geq \omega_\tau \left[\sum_i \gamma^i \phi_n^i - x_n\right] \end{array} \right\}, \text{ if } y_n \leq y_{n,\min}^{\text{cap}} \tag{2.12}$$

$$Z_n \geq 0, \quad \text{otherwise} \quad \forall n \tag{2.13}$$

Note that the lowest value of the departure capacity ($y_{\min}^{\text{cap}}$) is realized at the highest observed arrival count ($x_{\max}$) owing to the assumption of concavity for the departure capacity function.

The choice of quantile τ for the unhindered arrival capacity estimate is independent of its counterpart for the departure capacity function. The principle governing this choice in both cases is effective outlier elimination [31]. The choice is determined by iterating over a progression of quantiles descending from 100 percentile at chosen step sizes (set at 0.25 percentile in this study), until stable functional parameter estimates are obtained. In other words, the choice of τ was such that the quantile functions, be it the departure capacity function ($Q_\tau^{\text{dep}}(y|x)$) or the unhindered arrival capacity ($Q_\tau^{\text{arr}}(x)$), did not change significantly for a small change in the respective τ. This procedure is elaborated in Appendix II.

The statistical significance of each incremental vector of influencing factor (θ and ϕ) was ascertained through a hypothesis testing framework based upon the quantile likelihood ratio tests described in [47]. The details of this hypothesis testing framework are outlined in

Appendix I.

Lastly, the above formulation for the estimation of the departure capacity function inherently assumes concavity, complying with the a priori notion of convexity for capacity envelopes. Appendix III describes how the proposed estimation framework can be further adapted to statistically validate this convexity assumption. The key idea here is to relax the concavity constraint (2.7) in the above estimation formulation, and measure the improvement in the statistical fit of the quantile function. The details of this statistical testing framework are provided in Appendix III.

2.4.3   Comparison to prior approaches

As pointed out earlier, the chosen quantile (τ) for representing the capacity envelope components in the proposed methodology is a notional equivalent of the frequency filter in the method proposed by Gilbo [31]. However, in contrast to frequency-based filtering, a regression-based approach has the ability to quantify underlying factors that influence the capacity curve. This is done through the process of hypothesis testing where different functional specifications are compared using statistical metrics. The log-likelihood tests, as described in Koenker (2001) [46], were used for performing the hypothesis testing, and are elaborated in Appendix I. Each functional specification varies based on the set of factors considered in θ or γ. This means that, in addition to estimating capacity envelopes for each runway configuration, the proposed approach can potentially identify specific characteristics of the configuration (such as, the angle between the active arrival runways) that impact the capacity envelope in a statistically significant way.

Also, when it comes to assessing whether a given throughput observation is an outlier, the LP-based estimation framework will intrinsically incorporate the proximity of the observation to the estimated capacity envelope in addition to its observed frequency. This feature is enabled by the concavity constraints (2.7)-(2.8) enforced in the estimation formulation, and ensures that two of the candidate outlier rejection criteria suggested by Gilbo [31] are simultaneously accounted for. The beneficial effect of this feature is illustrated in Figure 2-5 below.

As demonstrated above, Gilbo's method [31] of frequency-based filtering would eliminate
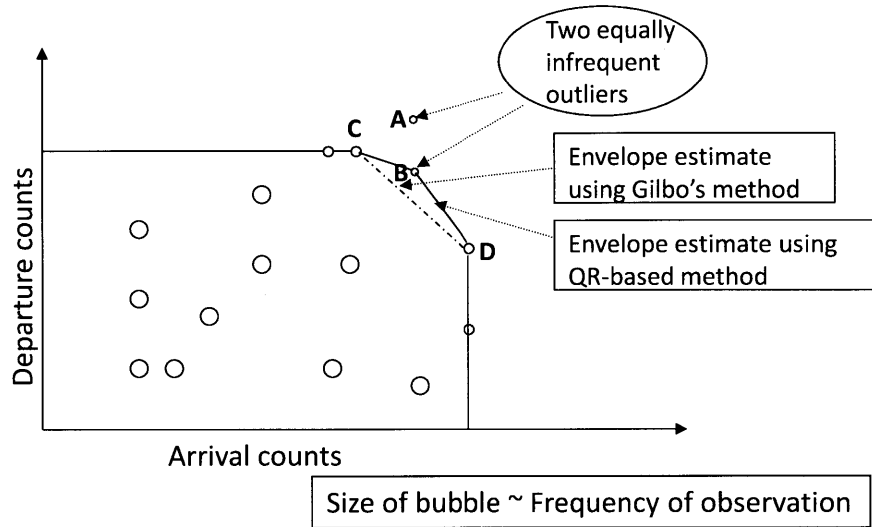
Figure 2-5: Outlier elimination: Proposed methodology vs Gilbo's methodology

observed throughput points A and B owing to their low frequency, while the quantile regression methodology would retain point B because of its conformity to the concave shape of the capacity envelope as dictated by the neighboring high-frequency observations C and D. The latter outcome is conceivably more desirable, as the credibility of point B is higher than that of A in this example due to the above-mentioned reasons.

We note that both the method proposed in this chapter and the approach proposed by Gilbo estimate the best-case capacity envelopes, that is, the capacity tradeoffs when the airport is at peak performance, with a favorable fleet mix, etc. However, airport (departure) capacity can alternatively be defined as the *average number of departures under sustained demand* [21]. Under this definition, the estimates of capacity envelopes can be quite different from the ones we obtain [70].

2.5 Case Study: New York Metroplex

The proposed formulations were applied to obtain capacity envelope estimates for the multi-airport system in NY consisting of three major airports: JFK, LGA and EWR. In the case of the NY system of airports, capacity tradeoffs exist between operations within each airport as well as across the airports due to the metroplex phenomenon discussed earlier. The focus

46

of the case study was to analyze all capacity tradeoffs within the NY airport system.

2.5.1 NY airport system overview

The three airports EWR, JFK and LGA are located in close proximity, and serve a considerable volume of operations simultaneously. The use of the shared airspace encompassing the three airports is coordinated by the TRACON operations in tandem with the local air traffic control (ATC) towers of the airports. The TRACON oversees the allocation of airspace to the arrivals and departures at the three NY airports, while the individual ATCs allocate runways and terminal airspace to the arrivals and departures serviced at the respective airports.

The three airports are equipped with 4, 3 and 2 runways respectively, and their relative alignments are shown in Figure 2-6. Note that the principal runways at the three airports are aligned with each other.

Figure 2-6: A map of the New York area, showing the approximate locations of the three core airports and their relative layouts. Note that the airport layouts are not to scale with the map. © Google. Image © 2009 DigitalGlobe. Image © 2009 Sanborn.

2.5.2 Capacity envelope estimation

The interaction between operations at different airports within the system was investigated using pairwise inter-airport capacity envelopes (for example, JFK arrivals vs. LGA departures, JFK departures vs. EWR departures, etc.). This is in addition to the intra-airport capacity envelopes relating arrival and departure capacities for each airport. Under this framework, the relationships among operational capacities at the three NY airports can be represented through 15 capacity envelopes: 3 intra-airport, and 3×4 inter-airport pairs.

It is assumed that for any given observation, the operational capacity of arrivals or departures at an airport was constrained by at most one of the remaining 5 operation types in the 3-airport Metroplex (departure or arrival at the same airport, and arrivals and departures at the other two). That is, at a given time, at most one of the 15 capacity envelopes is binding for a given operation type from a given airport in the NY Metroplex, be it LGA arrival, JFK departure or any other. This assumption implies that the intra-airport and inter-airport capacity envelopes can be estimated independently. The capacity for an operation type in the metroplex at any time would then be determined by the most restrictive pairwise envelope featuring it for that time interval.

The choice of the dependent and independent variable in each pairwise capacity envelope is determined by the relative congestion experienced at the airports, combined with the earlier observation that arrivals have greater priority than departures. When arrival operations from one NY airport are paired with departure operations from the same or another NY airport (e.g. LGA arrivals vs. JFK departures, EWR arrivals vs. EWR departures, etc.) the arrival is treated as the independent variable based on the above argument of operational prioritization. For inter-airport pairs featuring purely departure or arrival operations (e.g. LGA departures vs. JFK departures, EWR arrivals vs. JFK arrivals, etc.), the operation at the airport with traditionally higher congestion is chosen as the independent variable. The rationale behind this choice is that the traditionally busier airport is accorded preference by TRACON during conflicts in airspace usage.

From observed airport operational data for years 2005 and 2006 as well as historical anecdotes, it was inferred that JFK was the busiest of the NY airports, followed by LGA and

EWR in that order. This inference sets up the following precedence order among operations in the NY metroplex: JFK arrivals $\succeq$ LGA arrivals $\succeq$ EWR arrivals $\succeq$ JFK departures $\succeq$ LGA departures $\succeq$ EWR departures. For a capacity envelope involving a pair of operational types from the provided sequence, the operational type with higher precedence will serve as the independent variable.

2.5.3   Data sources for capacity estimation

The FAA's Aviation System Performance Metrics (ASPM) database provides records of flight activity at 77 of the major airports in the United States [25]. For each airport, for every 15-min interval, the database includes reports of the number of arrivals, the number of departures, prevailing weather conditions (Visual or Instrument Flight Rules, wind speed and direction), and the runway configuration used. The archived data for the three New York area airports were procured for this study. Overnight periods of operation (midnight-6AM) and the 15 minute intervals before and after runway configuration changes were filtered out from the analysis, due to the increased tendency for reporting errors during these periods. The estimation data set covered the years 2005 and 2006 for JFK and EWR, but was restricted to the year 2006 for LGA due to inconsistencies in the throughput reports during 2005. The linear programs for estimating the capacity envelopes were coded in AMPL [26], and solved using CPLEX [42] with the default primal-dual simplex method.

The following section describes the estimation results for the pairwise capacity profiles at the New York area airports, and discusses their implications. The presented capacity envelopes for each pairwise combination were finalized after statistically testing different explanatory factors using the hypothesis testing framework described in Appendix I.

2.6   Results

2.6.1   Intra-airport capacity tradeoffs

The influencing factors considered for the capacity tradeoffs within each airport include visibility (VFR or IFR), alignment of the arrival and departure runways (parallel or cross-

ing), and the number of additional runways for arrival or departure operations (beyond the primary runway). To provide insight into how these factors were incorporated into the estimation process in the form of θ and ϕ from Section 2.4.2, the finalized quantile specifications for the departure capacity function as well as unhindered arrival capacity at JFK are presented below.

Departure capacity function:

$$
\begin{aligned}
Q_\tau^{\text{dep}}(y_{\text{JFK}}|x_{\text{JFK}}) = \quad & \alpha + \alpha_k^{\text{prim}}\theta^{\text{prim}} + \alpha_k^{\text{ex-dep}}\theta^{\text{ex-dep}} + \alpha_k^{\text{ex-arr}}\theta^{\text{ex-arr}} + \alpha_k^{\text{ex-par}}\theta^{\text{ex-par}} + \\
& (\beta + \beta_k^{\text{prim}}\theta^{\text{prim}} + \beta_k^{\text{ex-dep}}\theta^{\text{ex-dep}} + \beta_k^{\text{ex-arr}}\theta^{\text{ex-arr}} + \beta_k^{\text{ex-par}}\theta^{\text{ex-par}})x_{\text{JFK}}, \\
& \qquad\qquad\qquad\qquad \text{for } (k-1) \leq x_{\text{JFK}} \leq k, \ \forall k \in \{1,...,20\}.
\end{aligned}
$$

Unhindered arrival capacity:

$$
Q_\tau^{\text{arr}}(x_{\text{JFK}}) = \gamma + \gamma^{\text{vis}}\phi^{\text{vis}} + \gamma^{\text{ex-par}}\phi^{\text{ex-par}} + \gamma^{\text{ex-conv}}\phi^{\text{ex-conv}}
$$

Variable definition

θ^{prim} :   Alignment of primary arrival and departure runways

     = { 1, for crossing runways

      2, for converging runways

      3, for parallel runways}

$\theta^{\text{ex-dep}}$:   Alignment of extra departure runway with arrival runway

     = { 0, if no extra departure runway, or if extra departure runway is the same as the arrival runway

      1, if extra departure runway crosses arrival runway

      2, if extra departure runway converges towards arrival runway without crossing

      3, if extra departure runway parallel to arrival runway}

$\theta^{\text{ex-arr}}$:   Alignment of extra arrival runway with departure runway

     = { 0, if no extra arrival runway or if extra arrival runway same as departure runway

      1, if extra arrival runway crosses departure runway

$$2, \text{ if extra arrival runway converges onto departure runway without}$$

crossing

$$3, \text{ if extra arrival runway parallel to departure runway}\}$$

$\theta^{\text{ex-par}}$: Incremental advantage of parallel runways pairings 13R | 13L or 31L | 31R over 22R | 22L or 4L | 4R

$$= \{ \ 1, \text{ if primary arrival|departure runways are 13R|13L or 31L|31R}$$

$$0, \text{ otherwise } \}$$

ϕ^{vis} : Visibility status

$$\{1, \text{ if VFR; } 0, \text{ if IFR}\}$$

$\phi^{\text{ex-par}}$: Additional parallel arrival runway

$$\{ \ 1, \text{ if additional arrival runway parallel to primary arrival runway}$$

$$0, \text{ otherwise } \}$$

$\phi^{\text{ex-conv}}$: Additional converging arrival runway

$$\{ \ 1, \text{ if extra arrival runway converges onto primary arrival runway}$$

$$0, \text{ otherwise } \}$$

The order of quantile (τ) yielding robust estimates for the departure capacity functions were found to be 99%ile, 99%ile and 99.5%ile for JFK, EWR and LGA respectively, while the equivalents for the unhindered arrival capacities were found to be 99.75%ile, 99.5%ile and 99.75%ile.

Figures 2-7, 2-8 and 2-9 illustrate the effect of the estimated influences on the capacity envelopes, under VFR and IFR conditions, for the three airports.

As shown in Figures 2-7, 2-8 and 2-9, visibility has a significant influence on the unhindered arrival capacities, but does not appear to have a noticable effect on departure capacities. The relative alignment of the primary runways used for arrivals and departures plays a critical role in determining the capacity envelope shapes. Figures 2-7 and 2-8 show that the area under the capacity envelope progressively increases as we go from mixed arrival/departure operations on a single runway, through separate arrival and departure runways that intersect or converge, to additional parallel runways, demonstrating the benefit of independent operations on runways. The use of an additional departure runway at JFK
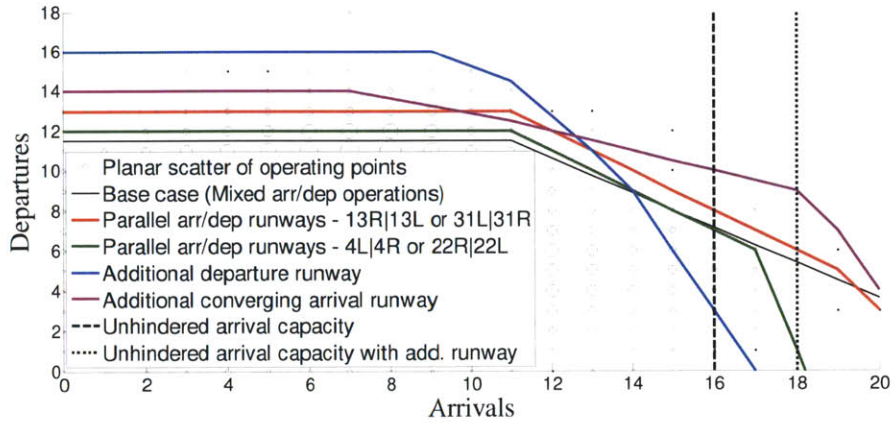
(a) VFR
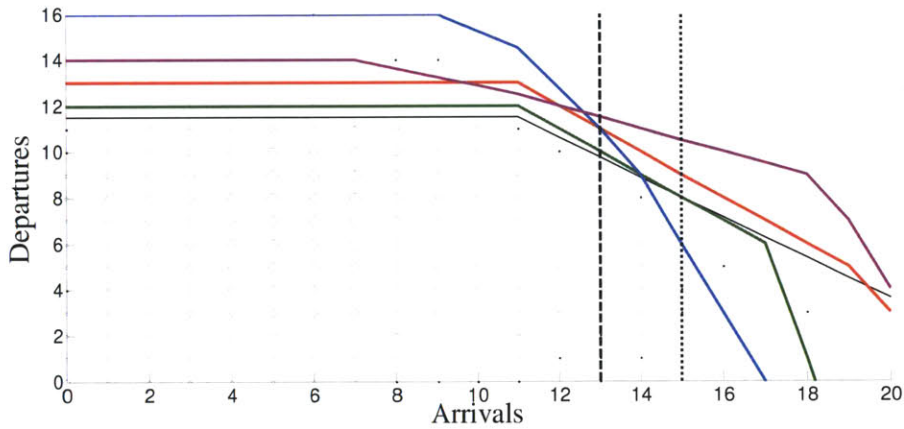
(b) IFR

Figure 2-7: JFK capacity envelopes @ 99 %ile

and EWR provides the most benefit in terms of departure capacity at low values of arrival rate. Figure 2-7 also shows that the use of an additional arrival runway at JFK flattens the slope of the tradeoff curve, indicating the effective redistribution of operations across the two runways, while also increasing the unhindered arrival capacity. It is also observed that the use of the parallel runway configurations (for example, 22R|22L or 4R|4L) at JFK results in a lower capacity as compared to their perpendicular counterparts (for example, 31R|31L or 13R|13L). This is possibly explained by the smaller distance between the former pair, resulting in a greater coupling of operations.
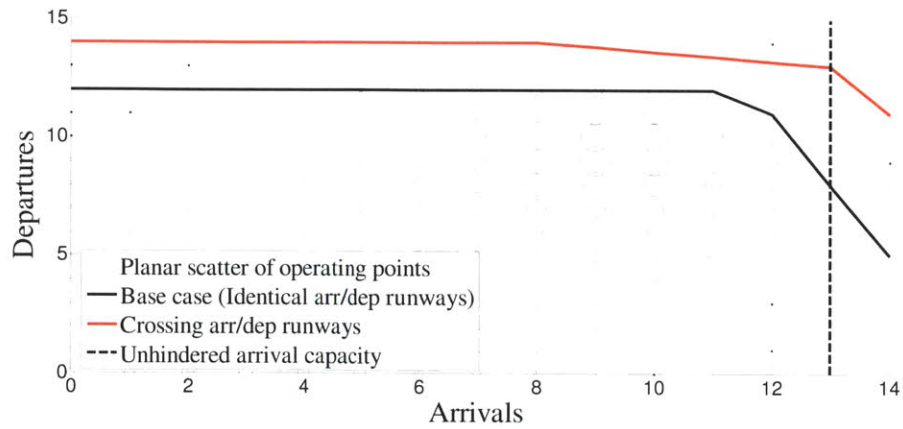
(a) VFR
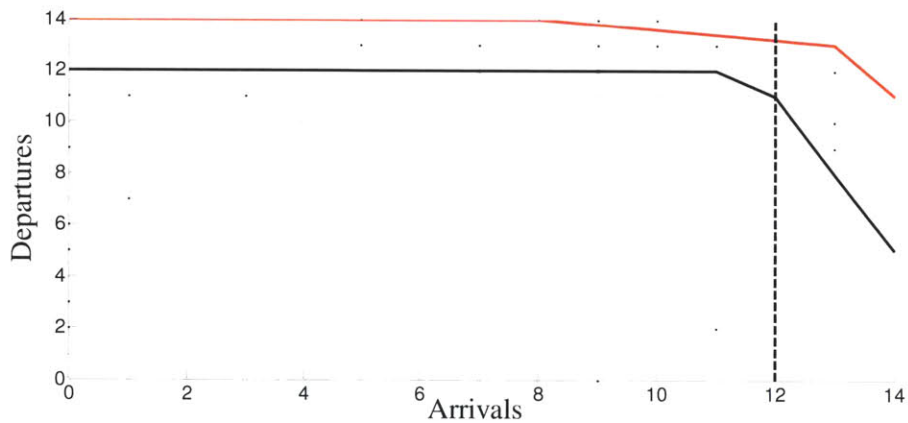
(b) IFR

Figure 2-8: LGA capacity envelopes @ 99.5 %ile

2.6.2 Inter-airport capacity tradeoffs

Since inter-airport interactions are expected to involve the airspace rather than the airport surface, the overlap between approach or departure paths is considered instead of the runway alignment attribute used for intra-airport capacity envelopes in Section 2.6.1. The approach and departure paths were approximated by two-dimensional conics with vertex angle of 30 deg extrapolated from the runway in the direction of operation, and binary terms were used to signify the intersection of these 2D conics. An example of the functional representation for inter-airport capacity envelopes is provided below for the pairing of JFK departures

53
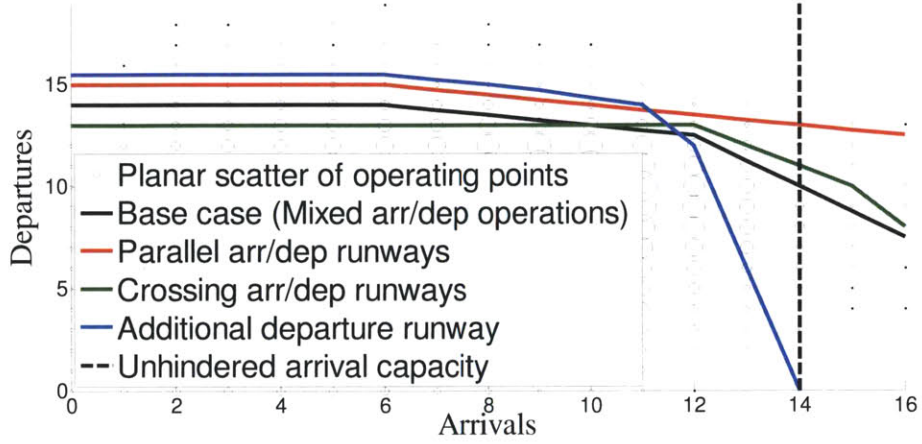
(a) VFR
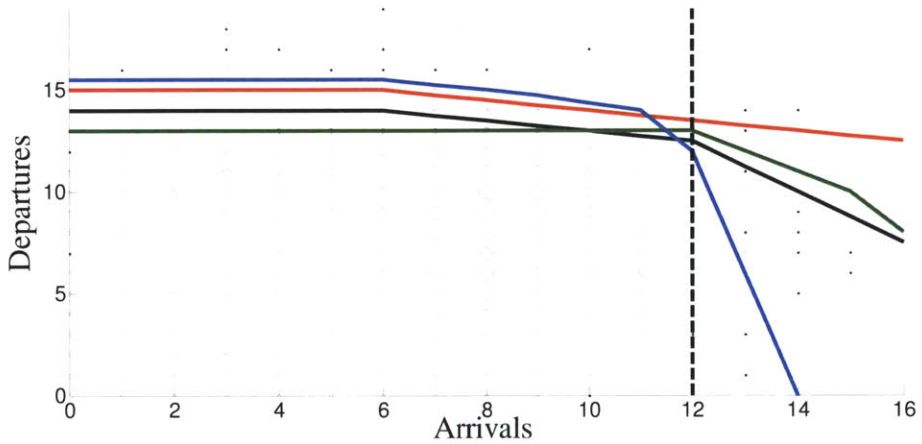
(b) IFR

Figure 2-9: EWR capacity envelopes @ 99 %ile

and EWR departures. Given the order of precedence for metroplex operations described in Section 2.5, JFK departure is chosen as the independent variable for this pairing.

Departure capacity function:

$$
\begin{aligned}
Q_\tau^{\mathrm{dep}}(y_{\mathrm{EWR}}|y_{\mathrm{JFK}}) = \quad & \alpha + \alpha_k^{\mathrm{vis}}\Theta^{\mathrm{vis}} + \alpha_k^{\mathrm{ex\_dep}}\Theta^{\mathrm{ex\_dep}} + \alpha_k^{\mathrm{int}}\Theta^{\mathrm{int}} + \\
& (\beta + \beta_k^{\mathrm{vis}}\Theta^{\mathrm{vis}} + \beta_k^{\mathrm{ex\_dep}}\Theta^{\mathrm{ex\_dep}} + \beta_k^{\mathrm{int}}\Theta^{\mathrm{int}})y_{\mathrm{JFK}}, \\
& \text{for } (k-1) \le y_{\mathrm{JFK}} \le k, \ \forall k \in \{1,...,16\} \quad (2.14)
\end{aligned}
$$

Variable definition

54

$\Theta^{\text{ex\_dep}}$:   Extra departure runway at EWR

$$= \begin{cases} 1, & \text{if extra departure runway,} \\ 0, & \text{otherwise} \end{cases}$$

Θ^{int} :   Intersection of projected takeoff paths for primary dep. runways at JFK and EWR

$$= \begin{cases} 1, & \text{if takeoff paths intersect} \\ 0, & \text{otherwise} \end{cases}$$

Representative inter-airport capacity envelopes for pairs of airports and arrival-departure operations, under different flight conditions are shown in Figures 2-10, 2-11 and 2-12.

From the inter-airport capacity envelopes, it is observed that capacity tradeoffs are prominent at higher throughput values than those seen in the intra-airport envelopes. This observation suggests that airport (ground) capacity is a more binding operational constraint than the capacity of the surrounding airspace. This system characteristic could be responsible for the negligible inter-airport tradeoffs estimated under IFR conditions and single runway configurations, since the airport throughputs are lower under these conditions. Figures 2-10, 2-11 and 2-12 illustrate this phenomenon for selected inter-airport operational pairs that exhibited tradeoffs close to the limits of their respective operational capacities. The approach path overlap attribute was not found to be statistically significant for any of these pairwise envelopes, possibly due to the limited operational range over which the tradeoff effects were found to be prominent.

2.6.3   Limitations of approach

This chapter developed a statistical framework for quantifying arrival-departure tradeoffs in a multi-airport system, and applied the framework to estimate airport capacity envelopes using 15-min throughput records in ASPM database. This section examines the estimated results in detail and highlights potential limitations of the approach that could be addressed in future extensions.
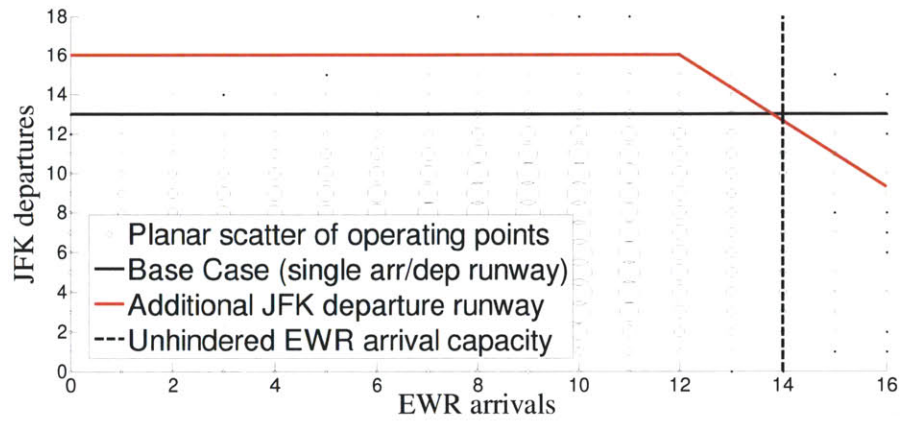
(a) VFR
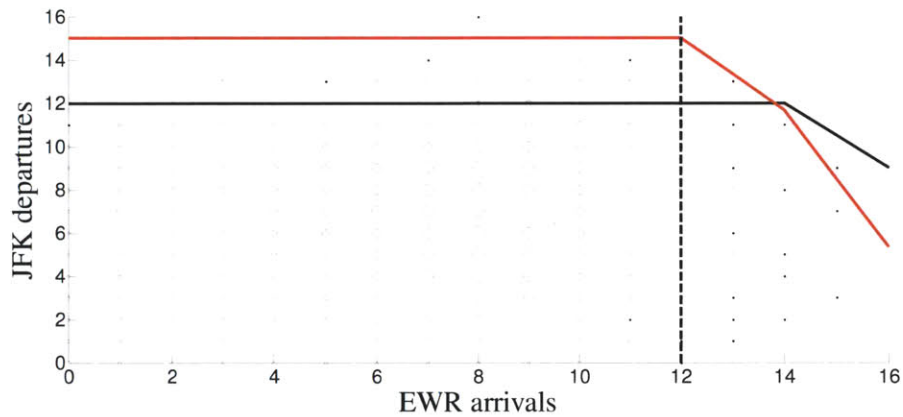
(b) IFR

Figure 2-10: JFK departure vs. EWR arrival capacity envelopes @ 99 %ile

2.6.3.1 ASPM data resolution

First, we consider the intra-airport envelopes for LGA (Figure 2-8 for τ=99.5%ile), and explore estimated results for different quantiles (τ). We focus on the single runway (mixed operations) case, since the estimated capacity envelope appears to be significantly large compared to the FAA Capacity Benchmark report's *optimal* runway configuration capacity estimate of 85 operations (41 arrivals, 44 departures) per hour [54].

Figure 2-13 depicts the capacity estimates for the single runway configuration for quantiles ranging from 99%ile to 50%ile. We notice that while the 99 %ile envelope is larger than the
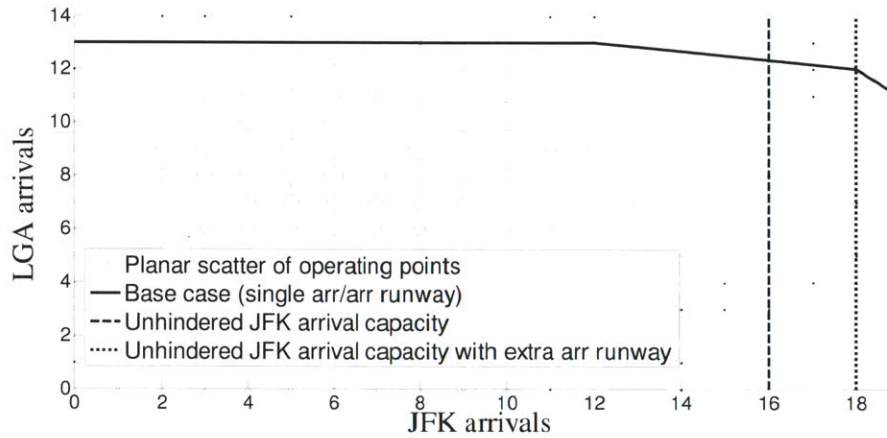
(a) VFR
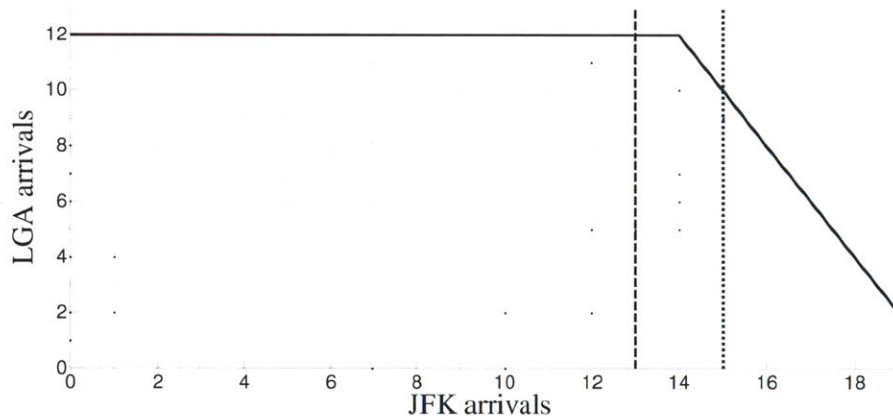
(b) IFR

Figure 2-11: LGA arrival vs. JFK arrival capacity envelopes @ 99.5 %ile

benchmark capacity for the optimal configuration, several of the other quantile estimates (as low as 75 %ile) have estimated capacities of 72-80 operations/hour for the single runway configuration. While this may seem surprising, Table 2.3 describes the frequencies of the fringe observations (points on and above the estimated envelope) in the arrival-departure trade-off region (arrival counts ≥ 10) for each of these estimates. We note that even after filtering out overnight observations, operating counts as high or higher were observed more than 160 times over two years. The selection of the right quantile is also a question of practical importance, and one that warrants further research.

As seen in Table 2.3, the fringe observations in the trade-off region for the estimated

57
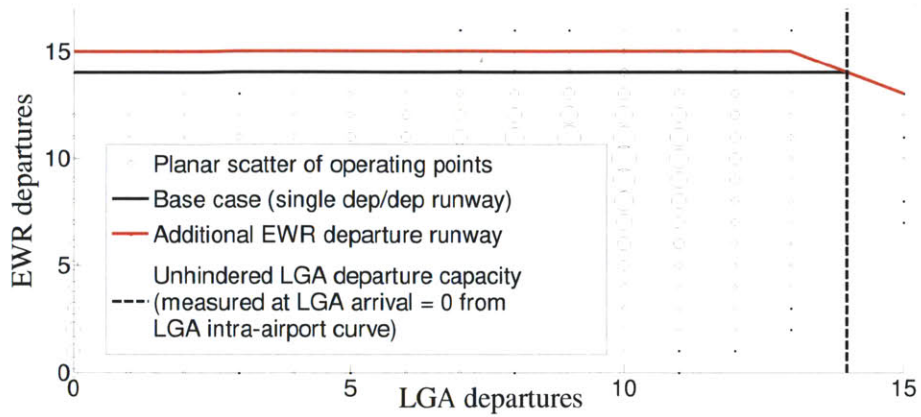
(a) VFR
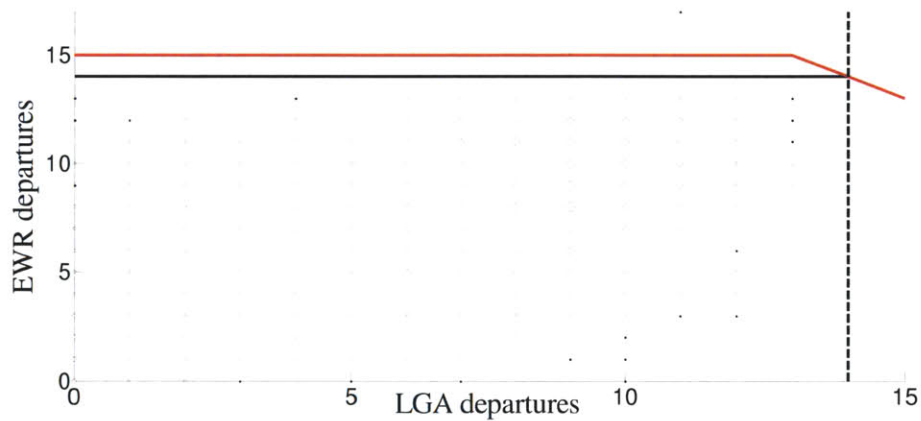
(b) IFR

Figure 2-12: EWR departure vs. LGA departure capacity envelopes @ 99.5 %ile

capacity envelopes at higher quantiles appear rather infrequently. However, due to the relatively limited number of total observations in this region (317), these points fall within the respective quantile range and are thus included within the respective envelope estimates. We note that the sparseness of total observations in this region limits the reliability of the estimates for the single runway configuration at LGA. Further investigation of this issue would require higher fidelity data sets, such as flight-specific records or surface surveillance data.
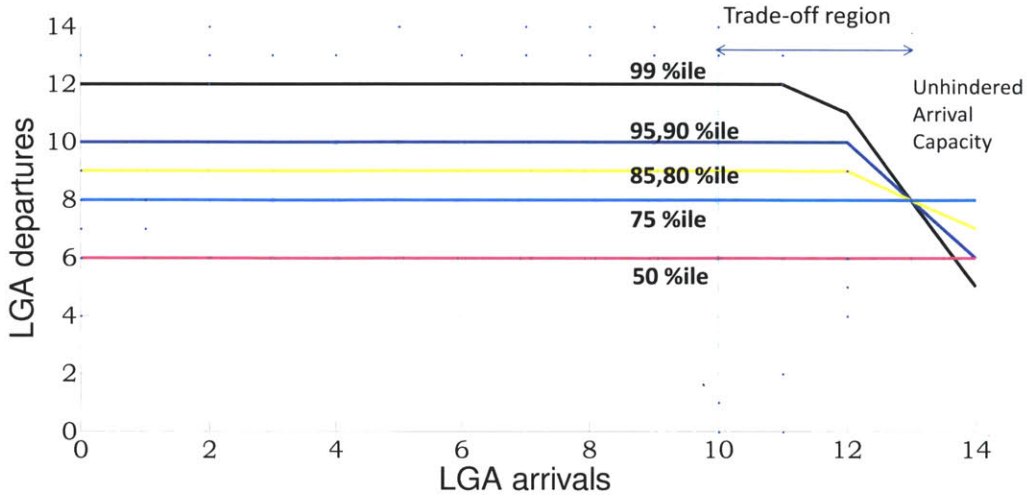
Figure 2-13: Capacity envelope estimates for various quantiles for single runway configuration at LGA.

τ %ile	No. of observations lying on capacity envelope	No. of observations above capacity envelope (Outliers)
99	13	5
95	37	22
90	37	22
85	45	54
80	44	54
75	75	89
50	54	192

Table 2.3: Frequency of fringe observations in trade-off region (Arrivals $\geq$ 10). Total observations in this region = 317.

2.6.3.2 Pairwise inter-airport tradeoffs

In addition to the above limitations of the estimation dataset and their impact on capacity envelope estimates, we also need to be mindful of the assumptions underlying the inter-airport capacity estimation while interpreting the corresponding estimates. As described in Section 2.5.2, the pair-wise inter-airport envelopes for the NY metroplex were estimated independently based on the assumption that no more than one inter-operational capacity constraint was binding for any given 15-min observation. In the estimation dataset for inter-airport envelopes, we do not explicitly separate observations for which the intra-airport envelopes were active. Hence, the resulting estimates could possibly be an underestimate of

59

the stand-alone inter-airport tradeoffs. In spite of this caveat, we notice in our final estimates that inter-airport tradeoffs come into effect at higher operational throughputs compared to the intra-airport tradeoffs.

2.6.3.3 Choice of dependent and independent variables

The choice of the dependent and independent variable for each inter-airport operational pair critically influences the obtained capacity envelope estimate. As mentioned in Section 2.5.2, after considering data from 2005 and 2006, JFK was assumed to be the busiest of the NY airports, followed by LGA and EWR. This assumption resulted in the following precedence order: JFK arrivals $\succeq$ LGA arrivals $\succeq$ EWR arrivals $\succeq$ JFK departures $\succeq$ LGA departures $\succeq$ EWR departures. We recognize, however, that this ordering of operations may change from year to year. It would therefore need to be revised depending upon the prevailing congestion patterns at the three NY airports for the chosen year of estimation.

In continuation of this chapter's theme of accurately estimating airport capacity, Chapter 3 focuses on airport operational dynamics including the processes of configuration selection and switching that have a direct bearing on the realized airport capacity. Together, the proposed models for characterizing and quantifying airport capacity supply the primary inputs for models of capacity allocation discussed in the later chapters.

Chapter 3

Characterization of Runway Configuration Dynamics

3.1 Introduction

The runway configuration is the subset of the runways at an airport used for arrivals and departures at any time. The focus is primarily on the process of runway configuration selection, a choice periodically made by airport authorities and driven by operational and regulatory considerations. As seen in Chapter 2, an airport's arrival and departure capacity at any time depend on the active runway configuration. Therefore, this chapter investigates the key airport operational processes that influence the attainable capacity.

Another important aspect of airport capacity dynamics is the transitional impact of configuration switching on airport capacity, given that the task of reconfiguring an airport is not instantaneous, and can cause disruptions to regular operations. This chapter develops statistical models to characterize configuration selection and configuration switch effects and uncover insightful trends on airport operations using recorded observations. Since the runway system is a critical bottleneck in airport operations, improved understanding of runway configuration dynamics can facilitate performance improvements and lead to system-wide benefits. Within the context of this research, models for configuration dynamics can help develop accurate forecasts for airport capacity that are subsequently used in the capacity allocation process. The rest of this chapter is divided into two parts: the first part presents a

model to describe the configuration selection process, while the second outlines an approach to estimate the duration and capacity impact of configuration switches.

3.2 Configuration Selection Model

Most major airports are equipped with multiple runways, but at any time, only a subset of these runways (and associated traffic directions) are selected to handle arrivals and departures. Authorities in the Air Traffic Control Tower (ATCT) consider many factors including weather (wind and visibility), predicted arrival and departure demand, environmental considerations such as noise abatement procedures, and coordination of flows with neighboring airports, in selecting the runway configuration at any time.

Figure 3-1 is a sketch of the airfield layout at LGA airport, which has two crossing runways aligned at 90° to each other and four operable runway orientations labelled 4, 13, 22 and 31 respectively. Some standard runway configurations utilized at LGA include 22|13,
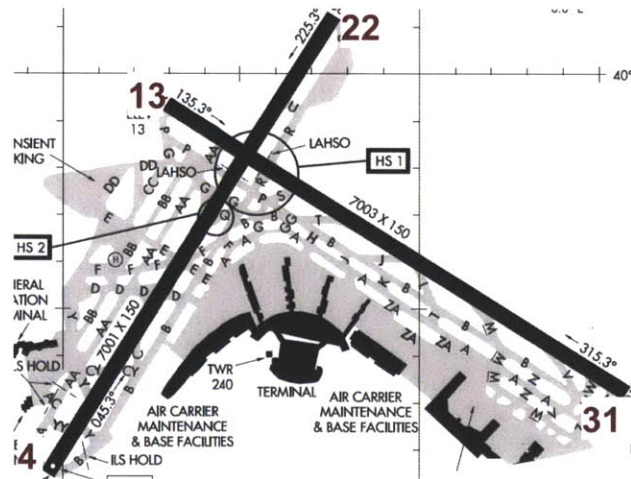
Figure 3-1: LGA runway layout

22|31, 31|4 etc., where the runway label on either side of the vertical bar identifies that arrival and departure runway respectively. While LGA authorities could choose to operate arrivals and departures on both runways, such an option might not always be required, possible or preferable under typical considerations. For example, the wind direction and speed at a given time might rule out the feasibility of operations on a particular runway from a safety standpoint. Concerns over noise mitigation could render a particular runway

orientation inoperable during specific hours of the day. Air traffic controllers thus adopt an ad hoc decision rule guided by above-mentioned considerations to select the configuration at a given time, and this rule could vary across airports.

Several past works have acknowledged the role of runway configuration selection in airport congestion management [29, 31, 35], and interest has thereby grown in the problem of configuration planning. Recent research has focused on the development of decision support systems that prescribe the optimal sequencing of runway configurations, assuming knowledge of their respective capacities, expected airport demand, and prevailing operating conditions influencing configuration feasibility [27, 51]. This research takes a complementary approach to the problem of configuration selection and addresses the task of describing how controllers select configurations as opposed to determining how they should be. The configuration selection processed is modeled using a discrete choice framework [10]. The relative significance of the factors governing configuration selection are captured within configuration utility functions and estimated using archived observations through the likelihood maximization method. The discrete-choice framework enables the estimation of the relationship between attributes of alternatives and their favorability as evaluated by a choice-maker. This modeling approach has been successfully applied to several applications involving choice-making among a set of discrete alternatives, such as consumer purchases [10], driver lane-changing behavior [1, 2], etc. In this application, the airport traffic controllers serve as the choice-maker and the various feasible configurations constitute the discrete alternatives at each time step. The discrete choice model can be estimated using observed configuration choices under recorded conditions, and the estimated model can be used to predict future configuration choices made by controllers in response to evolving weather and demand conditions.

The next section provides an overview of a typical discrete-choice model application, starting with the model framework and encompassing the tasks of model estimation and validation. This discussion is tailored to the specific problem of configuration selection. Results from the application of this modeling approach to LaGuardia (LGA) and Newark (EWR) airports are used to demonstrate its ability to predict the runway configuration, given the state of the system in terms of wind, visibility, demand, etc.

63

3.3 Discrete Choice Methodology

3.3.1 Conceptual framework

Discrete-choice analysis [10] considers problems in which a decision-maker needs to select one option from a finite set of alternatives. It is assumed that the decision-maker chooses the solution that maximizes a utility function that depends on several influencing factors (known as attributes, and denoted X). The utility function for each alternative is modeled as the sum of an observed component V (which is a linear combination of the influencing factors) and an unobserved component ϵ represented through error terms. Consider a particular choice observation from a sample set, arbitarily labeled the nth observation. Suppose C_n is the set of alternatives available for the choice maker in this observation. Then, the utility of alternative $c_i \in C_n$ for this choice process is given by

$$V_{in} = \alpha + \beta \cdot X_{in} \tag{3.1}$$

$$U_{in} = V_{in} + \epsilon_{in}, \tag{3.2}$$

Equation (3.1) provides the expression for the observed utility component, and reflects the assumption that the utilities are linear functions of the attributes X_{in}. Equation (3.2) acknowledges the presence of errors in the utility perceived by the choice maker. These errors represent factors that are not explicitly observed or included amongst X_{in}. The choice-maker is assumed to make a rationale choice and select the alternative $c_j \in C_n$ with maximum utility U_{jn}.

$$j = \underset{i:c_i \in C_n}{\operatorname{argmax}} U_{in}. \tag{3.3}$$

The unobserved error term ϵ_{in} is assumed to follow a probabilistic distribution, thereby rendering the choice process a stochastic event with each alternative having a specific selection probability. The probabilistic distribution assumed for the error terms ϵ_{in} determines the analytical relation between alternative selection probabilities and the observed component of the utility functions, and hence the type of discrete choice model. When one assumes complete independence in error terms across all alternatives and choice observations, and that

the error terms are identically Gumbel distributed, the multinomial logit (MNL) model [10] is obtained. The MNL model is a popular choice in many applications due to its analytical tractability, and yields the alternative selection probability expression given by

$$P(c_i|C_n) = \frac{e^{V_{in}}}{\sum_{j:c_j \in C_n} e^{V_{jn}}}. \qquad (3.4)$$

In other words, equation (3.4) provides the probability that the selection for the nth choice process was alternative c_i, given that the set of feasible alternatives for this process was C_n. Note that as the observed component of the utility for alternative c_i (given by V_{in}) increases relative to the equivalent values of the other alternatives, so does the probability of selecting c_i.

The assumption of independent error terms across all alternatives, as adopted in the MNL model, is potentially too restrictive in the context of runway configuration selection. For instance, consider two feasible configurations that contain a common arrival (or departure) runway. This common runway might contribute identical unobserved effects to the configuration utilities, rendering their error terms correlated. To mitigate this shortcoming, advanced versions like the Nested Logit (NL) and Cross-Nested Logit (CNL) models [10] are considered for this discrete choice application. These model structures permit error correlation within specified subsets of alternatives as illustrated in the nested frameworks shown in Figure 3-2. The illustration pertains to a choice example featuring four alternatives {alt1, alt2, alt3 and alt4}. The alternatives are grouped into two nests in (a) an exclusive manner (NL representation), and b) an overlapping manner (CNL representation) with alt2 shared between the two nests in the later structure. Note that some nests can be singletons. In the NL structure, alternatives alt1 and alt2 would have a common component in their error terms, and likewise for alternatives alt3 and alt4, but alternatives across nests, like alt2 and alt 3, would have independent errors. In the CNL structure, alt2 would have a common component of error with alt1, as well as with alt3 and alt4.

The expressions for alternative probabilities for the NL and CNL models, and their comparisons with the MNL model are described in [10]. For example, the selection probability
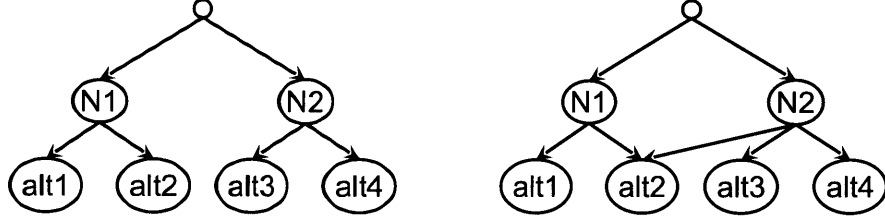
Figure 3-2: (a) NL model framework; (b) CNL model framework.

for alternative alt1 in the NL model (Figure 3-2 (a)) is given by

$$P(\text{alt1}|\{\text{alt1}, \text{alt2}, \text{alt3}, \text{alt4}\}) = P(\text{alt1}|N1)P(N1|\{N1, N2\}), \tag{3.5}$$

where $P(\text{alt1}|N1) = \frac{e^{\mu_{N1} * V_{\text{alt1}}}}{\sum_{j:c_j \in \{\text{alt1},\text{alt2}\}} e^{\mu_{N1} * V_j}}$, $P(N1|\{N1, N2\}) = \frac{e^{V_{N1}}}{e^{V_{N1}} + e^{V_{N2}}}$,

$V_{N1} = \frac{1}{\mu_{N1}} * \log \sum_{j:c_j \in \{\text{alt1},\text{alt2}\}} e^{\mu_{N1} * V_j}$, and similarly for V_{N2}.

Here, the scale parameters μ_{N1} and μ_{N2} provide a measure of the magnitude of error correlation among alternatives within nests N1 and N2 respectively. For the airport configuration choice problem, all three models (MNL, NL and CNL) were applied and investigated through appropriate statistical tests.

3.3.2 Estimation framework

The parameters of a discrete choice model are the coefficients of the observed influencing factors X_{in} on the alternative utilities U_{in} (α, β in Equation (3.2)), along with the scale parameters like μ for NL and CNL model structures. These are estimated using the maximum-likelihood approach. The likelihood of a given choice observation is simply the probability of selecting the observed choice given the values of the model parameters (α, β, μ) and influencing factors (X_{in}). The likelihood function for an entire dataset of choice observations (say, over N choice instances) is the joint probability of observing the sequence of choice decisions recorded, or in other words

$$\mathcal{L}(\alpha, \beta, \mu) = P((c_1|C_1) \bigcap .... \bigcap (c_N|C_N)|\alpha, \beta, \mu, X) \tag{3.6}$$

where c_i is the selected alternative, and C_i is the set of available alternatives for i^{th} observation, $i \in 1, 2, .., N$. In the configuration selection setting, a choice process is assumed to occur every recorded time step. In the estimation data set discussed for the case studies, each record spans a 15 minute time interval.

A typical assumption made in the estimation step is that the choice observations (at each time) are conditionally independent given the values of the explanatory factors X_{in}. This allows the likelihood function to be expressed as the product of the likelihood of individual choice observations from Equation (3.6).

$$\mathcal{L}(\alpha, \beta, \mu) \quad = \quad \prod_{i=1}^{N} P(c_i|C_i) \tag{3.7}$$

where $P(c_i|C_i)$ is given by Equation (3.4) for the MNL model or Equation (3.5) for the NL model.

The parameter estimates $(\widehat{\alpha}, \widehat{\beta}, \widehat{\mu})$ are those that maximize this likelihood:

$$(\widehat{\alpha}, \widehat{\beta}, \widehat{\mu}) = \underset{\alpha, \beta, \mu}{\mathrm{argmax}}\, \mathcal{L}(\alpha, \beta, \mu). \tag{3.8}$$

Note that the scale parameter μ is only estimated for NL and CNL models.

Likelihood-maximization is a nonlinear optimization problem. This study uses BIOGEME ([14]), a freeware package that specializes in estimating discrete-choice models through customized in-built algorithms.

3.3.3 Model specification and structure development

Model specification refers to the exact functional form of the systematic utility component V_{in}, comprising of the observed influencing factors X_{in}. The specification is developed through iterative investigation of candidate factors affecting the choice behavior. Standard hypothesis testing procedures help assess the statistical significance of every new factor considered. Likelihood-ratio test for nested hypothesis testing [10], and Cox composite model test for non-nested hypothesis testing [22, 19] are the two test designs used in this study to develop configuration utility specifications. The structure of a discrete choice logit model

refs to the particular correlation structure adopted for the alternative error terms ϵ_{in}. As mentioned earlier, MNL, NL and CNL models were all considered in this study. Established hypotheses tests (Hausman-McFadden test [38]) that help ascertain the statistical validity of structural enhancements offered by the NL or CNL model over the MNL model were used in this study.

3.4 Model Validation

The final step in any empirical model-building process is the evaluation of its predictive capabilities in comparison to a different, typically simpler, model that serves as the baseline framework. Both the proposed and baseline models are applied upon a validation dataset, using parameters estimated from a common training dataset, and their predicted probabilities are assessed, through well-defined metrics, for their proximity to the actual observed choices in the validation dataset. The definition of the baseline model is critical to the outcome of the validation task. This study adopts a probabilistic model depicting configuration selection as a Markovian transition process to be the baseline model. Such a model was proposed in [61], and differs from the discrete choice framework in the crucial sense that it does not explicitly model influencing factors X_{in} driving configuration selection.

The following section presents the details of the application of the proposed technique to the configuration selection process at LGA and EWR airports, as well as the associated results and inferences.

3.5 Case Study: LGA and EWR Airports

3.5.1 Training data set

The training data set consisted of the 15-minute aggregate ASPM records for the year 2006, which provide for each 15 minute interval, the chosen configuration as well as other prevailing airport conditions such as weather, wind speed and direction, demand, etc. Configuration selection is assumed to occur at every 15-min interval. Operational data for hours from 12 midnight to 6 am were excluded from the data set, since reporting during these periods

is more prone to errors. Feasible configurations for each time period were determined by the set of runways that did not exceed the FAA-specified safety thresholds for tail-winds (5 kn) and cross-winds (20 kn) [21]. Observations featuring operation of infeasible runway configurations (most likely reporting errors) were also excluded from the data set.

3.5.2 Candidate influencing factors

There are several factors that potentially influence the choice of configuration (from among the feasible options) in any time period. Presented below are those that are explicitly included in the utility functions of the discrete choice model.

Inertia: Configuration changes are a fairly involved procedure, require extensive coordination among the different airport stakeholders, and are thought to cause a loss in airport throughput [27, 51]. The latter aspect is exclusively studied in the second part of this chapter. For these reasons, the configuration from the previous time interval is likely to be favored pending other considerations, and its utility is therefore expected to be higher on account of this inertial factor relative to other alternatives.

Head-wind speeds: It is hypothesized that higher head-wind speeds are favorable for both arrival and departure operations, and therefore increase the utilities of the respective configurations. In this study, a combination of current and forecasted wind conditions are used as the measure of this influencing factor, given that controllers are likely to consider future conditions when planning configuration changes. In the absence of information on the actual forecast used by airport planners, the observed wind speeds over the immediate future of every time period is used as a proxy.

Arrival/departure demand: During periods of significantly high total (arrival + departure) demand, a high-capacity configuration is likely to be favored. The configuration-specific capacity envelopes obtained in the previous chapter are used to define this factor in the utility function.

Noise abatement procedures: In accordance to FAA procedures, certain runway configurations are to be avoided during applicable time periods. The Standard Operating

Procedures (SOPs) for the NY airports identify the overnight hours (10pm-7am) for activating the noise mitigation measures, and time-specific variables are accordingly defined for the configuration utilities in this study.

Configuration switch proximity: Configuration changes require increased coordination among airport elements, and disrupt the flow of aircraft on the surface. Authorities might therefore be inclined to minimize the level of effort involved. For example, a configuration change that only requires the addition of a departure or arrival runway may be easier to implement than a change that needs to change the direction of arrival flows entirely. In this study, the type and magnitude of the change is equated to the incident angles between the respective arrival and departure runways of the preceding and succeeding configurations. Using this representation, six distinct possible switch types are defined and their relative preferability is studied through appropriate categorical variables. For example, the configuration change which results in a 90° reorientation of the arrival runway and a 180° reorientation of the departure runway is denoted as switch category (90, 180) and applied to all configurations that will require such a change from the configuration used in the previous time interval.

Inter-airport coordination: In multi-airport terminal-areas such as New York, arrival and departure flows into the different airports must be coordinated. This coordination is handled by the NY TRACON in conjunction with the local ATCs of the airports, and can occasionally dictate configuration choices. Given this understanding, the effect of JFK's configuration on the concurrent choices for LGA/EWR is investigated. The nature of inter-airport configuration coordination depends upon the interactions between the respective runway orientations. Categorical variables representing these interactions between distinct pairs of runways at JFK and at LGA/EWR are defined for each configuration at LGA/EWR. Since airport authorities follow runway-specific airspace routes for landing and takeoff operations, the existing interactions among the routes from every pair of runways from the two neighboring airports can be estimated through this set of variables.

3.5.3 Estimation of discrete-choice models and utility functions

As explained in the methodological overview, the utility specifications and error structures were developed and statistically verified through a sequence of tests. The details of the finalized models are discussed below.

3.5.3.1 LGA results

The training data set had a total of 17,455 choice observations post-filtering (i.e., data from 17,455 15-min time periods), featuring a total of 10 distinct configuration alternatives. The final model has a NL structure with two alternative nests, grouping configurations with arrival runways 4 and 13 respectively as illustrated in Figure 3-3. The other configurations are modeled as singleton nests. For an understanding of the geometric alignment of the labels configurations, refer the LGA runway layout in Figure 3-1.
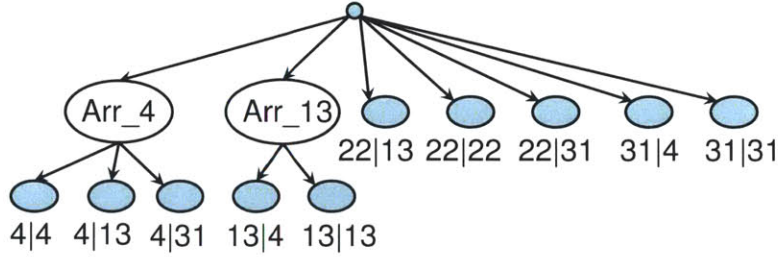
Figure 3-3: Estimated NL structure for LGA configuration selection (for year 2006).

The finalized configuration utility functions for LGA is as provided below.

$$
\begin{aligned}
V_{it} = {} & \alpha_i + \beta^i_{\text{vis}}\text{vis}_t + \beta_{\text{inertia}}(\text{choice}_{t-1} = i) \\
& + \sum_{k \in \{\text{arr,dep}\}} \beta^k_{\text{wind}}\text{wind}^k_{it} + \sum_{k=1}^{5} \beta^k_{\text{switch}}(\text{switch}(i, \text{choice}_{t-1}) = k) \\
& + \sum_{k \in \{\text{mor,eve}\}} \sum_{p \in \{4,13,22,31\}} \beta^k_p(\text{LGA\_run}^{\text{arr}}_i = k)\text{time}_k \\
& + \beta^{\text{dem}}_{\text{cross}}\text{dem\_ind}^t(\text{type}_i = \text{``cross''}) \\
& + \sum_{l \in \{\text{arr,dep}\}} \sum_{p \in \{4,13,22,31\}} \sum_{k \in \{4,13,22,31\}} \beta^l_{k,p}(\text{LGA\_run}^l_i = k)(\text{JFK\_run}^l_t = p)
\end{aligned}
\tag{3.9}
$$

71

Notation for utility variables:

choice_t : Configuration observed at time t

vis_t : $\begin{cases} 1, & \text{if VFR at time } t, \\ 0, & \text{otherwise} \end{cases}$

wind_{it}^{k} : Headwind speed along k^{th} runway (arr or dep) of configuration i at time t

$\text{switch}(i, j)$: Switch category between configurations i and j, defined based on angular reorientation between respective arrival and departure runways

$= \begin{cases} 1, & \text{if } (0, 90) \text{ or } (90, 0) \\ 2, & \text{if } (90, 90) \\ 3, & \text{if } (0, 180) \text{ or } (180, 0) \\ 4, & \text{if } (90, 180) \text{ or } (180, 90) \\ 5, & \text{if } (180, 180) \end{cases}$

$\text{LGA\_run}_i^{\text{arr}}$: Arrival runway of configuration i

$\text{LGA\_run}_i^{\text{dep}}$: Departure runway of configuration i

$\text{JFK\_run}_t^{\text{arr}}$: Arrival runway operated at JFK for time t

$\text{JFK\_run}_t^{\text{dep}}$: Departure runway operated at JFK for time t

time_k : $\begin{cases} 1, & \text{if } (\text{hr}(t) \in [6, 8]) \& (k = \text{morn}) \text{ or } (\text{hr}(t) \in [22, 24]) \& (k = \text{eve}) \\ 0, & \text{otherwise} \end{cases}$

type_i : Capacity class configuration i belongs to. "Mixed" or "Crossing" arrival/departure runways

dem_ind^t : Demand indicator variable

$= \begin{cases} 1, & \text{if arrival} + \text{departure demand} \geq 10 \\ 0, & \text{otherwise} \end{cases}$

The estimated results for the coefficients of the utility variables are tabulated in Table 3.2, along with the corresponding t-statistic in parenthesis. When the absolute value

72

1. Inertial	
β_{inertia}	+5.1 (68.8)

2. Visibility	
$\beta_{\text{vis}}^{13|13}$	+1.53 (2.72)
$\beta_{\text{vis}}^{22|22}$	+1.83 (2.79)
$\beta_{\text{vis}}^{31|31}$	+1.71 (2.2)
$\beta_{\text{vis}}^{4|13}$	+1.15 (3.42)
$\beta_{\text{vis}}^{13|4}$	+0.83 (1.47)
$\beta_{\text{vis}}^{22|13}$	+0.8 (2.1)
$\beta_{\text{vis}}^{22|31}$	+1.35 (2.71)
$\beta_{\text{vis}}^{31|4}$	+1.54 (4.3)

3. Headwind speed	
$\beta_{\text{wind}}^{\text{arr}}$	0.044 (8.2)
$\beta_{\text{wind}}^{\text{arr}}$	0.029 (5.0)

4. Demand	
$\beta_{\text{cross}}^{\text{dem}}$	1.68 (8.9)

5. Noise abatement	
β_{31}^{mor}	1.32 (7.4)
β_4^{eve}	0.93 (3.8)
β_{31}^{eve}	-0.29 (-1.3)

6. Switch proximity	
β_{switch}^4	-1.78 (-7.53)
β_{switch}^5	-2.2 (-4.4)

7. Coordination with JFK	
$\beta_{4,4}^{\text{dep}}$	0.44 (1.75)
$\beta_{13,13}^{\text{dep}}$	-0.404 (-1.37)
$\beta_{13,22}^{\text{dep}}$	-0.588 (-1.69)
$\beta_{13,31}^{\text{dep}}$	-1.05 (-3.55)
$\beta_{22,31}^{\text{arr}}$	-1.14 (-5.83)
$\beta_{31,4}^{\text{arr}}$	-0.4 (-1.7)
$\beta_{31,22}^{\text{dep}}$	0.959 (4.1)

8. Nest scale parameters	
μ_4^{arr}	1.1 (1.8)
μ_{13}^{arr}	1.65 (1.75)

Table 3.2: Estimation results for configuration selection at LGA.

of the t-statistic exceeds 1.96, the estimate of that parameter can be deemed statistically significant. As can be observed, the a priori hypotheses made in Section 3.5.2 are corroborated by the estimation results in the case of inertial effects and headwind speeds. While the estimates for the switch category and the JFK configuration coordination variables are hard to interpret due to less a priori understanding, a comparison with the corresponding estimates from an independent data set exhibits consistency in the configuration preferences.

3.5.3.2 Illustration of implied configuration selection probabilities using hypothetical values for utility variables

The bar plots in Figure 3-4 depict how the estimated coefficients translate to configuration selection probabilities given hypothetical values for the utility variables. The discussion is restricted to prominent runway configurations: 4|4, 31|31, 4|13, 22|13, 22|31 and 31|4. Hypothetical scenarios are created for illustrating the tradeoffs between switch proximity, wind favorability and operational capacity as influencing factors for configuration selection. VFR

operating conditions are assumed, and the simultaneous configuration at JFK is assumed to be its most prominent (31R|31L). It is also assumed that noise abatement regulations do not apply. Within this set of conditions, two demand scenarios are considered, low ($\text{dem\_ind}^t = 0$) and high ($\text{dem\_ind}^t = 1$). Assuming that the current runway configuration at LGA is 31|31, the relative selection probabilities of all the prominent configurations are examined for 20 kn wind speeds, with directions such that there are headwinds along runways 31, 22 and 13 respectively. Note that the probabilities presented for non-incumbent config-
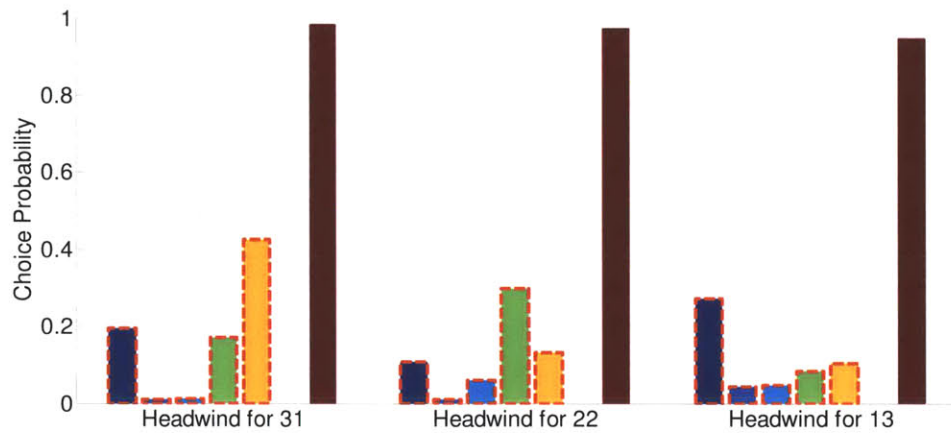
(a) Low Demand Scenario
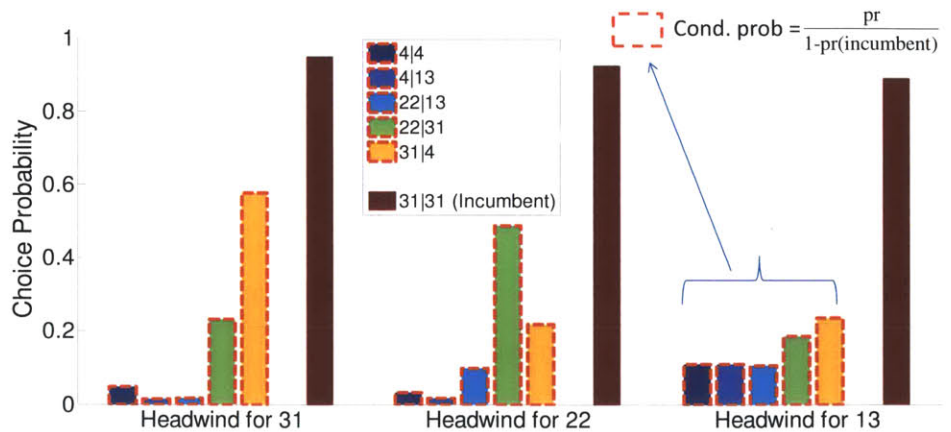
(b) High Demand Scenario

Figure 3-4: Relative configuration selection probabilities at LGA for described hypothetical scenarios.

urations are measured relative to each other (i.e, conditioned on a change in configuration) to facilitate comparison. The absolute selection probability for the incumbent is also shown next to these relative probabilities. For the low demand scenario (Fig. 3-4 (a)), it is observed that configurations with headwinds are typically favored among the non-incumbents, with the exception occurring when the wind blows along runway 13. In this scenario, although the non-incumbent configurations 4|13 and 22|13 have headwinds for departures, they would both require a less favorable switch (type 4) from the incumbent configuration 31|31, which reduces their desirability. In addition, the inertia effects ensure that the incumbent configuration (31|31) has a high probability of being retained for all three wind directions, although this probability progressively reduces as wind directions become less favorable.

For the high demand scenario (Figure 3-4 (b)), configurations with crossing runways (4|13, 22|13, 22|31 and 31|4) dominate among the non-incumbents, while the retention probability for the incumbent also comparatively reduces, highlighting the increased importance of higher configuration capacity over other considerations such as switch proximity.

3.5.3.3 EWR results

The training dataset had a total of 22,792 choice observations post filtering, featuring a total of 20 distinct configuration alternatives. The final model has a nested logit structure with one nest for a well-defined subset of alternatives as depicted in Figure 3-5. The nest gathers all EWR configuration alternatives with an additional arrival runway. The implication of this nesting is that configurations with an additional arrival runway share commonalities in terms of the unobserved factors influencing their preferences.
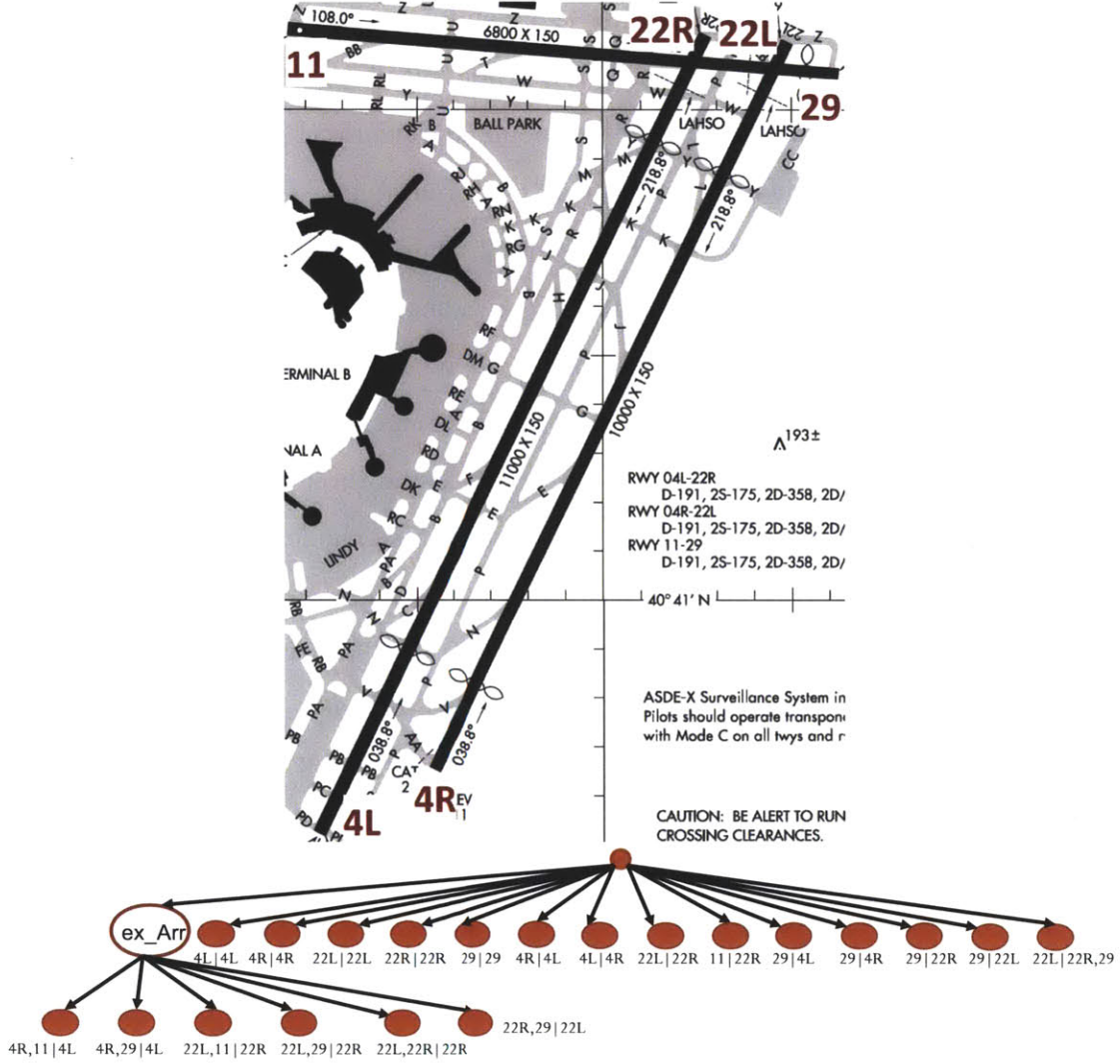
Figure 3-5: Layout of EWR, along with the estimated NL structure for EWR configuration selection (for year 2006).

The finalized configuration utility function for EWR is as shown below.

$$
\begin{aligned}
V_{it} = \quad & \alpha_i + \beta^i_{\text{vis}}\text{vis}_t + \beta_{\text{inertia}}(\text{choice}_{t-1} = i) \\
& + \sum_{j\in\{\text{prim,ex}\}} \sum_{k\in\{\text{arr,dep}\}} \beta^{j,k}_{\text{wind}}\text{wind}^{j,k}_{it} + \sum_{k=1}^{5} \beta^k_{\text{switch}}(\text{switch}(i,\text{choice}_{t-1}) = k) \\
& + \sum_{k\in\{\text{mor,eve}\}} \sum_{p\in\{4,11,22,29\}} \beta^k_p(\text{EWR\_run}^{\text{arr}}_i = k)\text{time}_k \\
& + \sum_{k\in\{\text{par,ex\_arr,ex\_dep}\}} \beta^{\text{dem}}_k\text{dem\_ind}^t_k(\text{type}_i = k) \\
& + \sum_{l\in\{\text{arr,dep}\}} \sum_{p\in\{4,13,22,31\}} \sum_{k\in\{4,11,22,29\}} \beta^l_{k,p}(\text{EWR\_run}^l_i = k)(\text{JFK\_run}^l_t = p)
\end{aligned}
\tag{3.10}
$$

76

Extended Notation (in addition to that described for LGA):

prim : Shorthand for primary runway (for arrival or departure) of a configuration

ex : Shorthand for extra runway (for arrival or departure) of a configuration

type_i : Capacity class configuration that i belongs to.

    "par" : Parallel arr/dep runways

    "ex_arr" : Extra arrival runway

    "ex_dep" : Extra departure runway

dem_ind_k^t : Demand indicator variable implying need for capacity class k

$\text{dem\_ind}_{\text{par}}^t$: $\begin{cases} 1, & \text{if arrival + departure demand} \geq \text{capacity of crossing} \\ & \text{runway configuration} \\ 0, & \text{otherwise} \end{cases}$

$\text{dem\_ind}_{\text{ex\_arr}}^t$: $\begin{cases} 1, & \text{if arrival demand} \geq \text{unhindered arrival capacity} \\ & \text{without additional runway} \\ 0, & \text{otherwise} \end{cases}$

$\text{dem\_ind}_{\text{ex\_dep}}^t$: $\begin{cases} 1, & \text{if arrival + departure demand} \geq \text{capacity of parallel} \\ & \text{runway configuration without additional runway} \\ 0, & \text{otherwise} \end{cases}$

The estimated results for utility coefficients are tabulated in Table 3.4, along with the corresponding t-statistics in parenthesis. It can be noted once again that most parameters are statistically significant.

Once again, the stated a priori hypotheses for inertia and wind effects are largely substantiated by the estimation results. Additional wind speed coefficients are introduced to capture effects on the supplementary (extra) runways independent of the primary runways for configurations with more than one arrival or departure runway. Also, the estimates for the switch category and the JFK configuration coordination variables were cross verified with those obtained for year 2007 to assess their credibility. The estimates for these variable types

1. Inertial	
β_{inertia}	+4.82 (66.6)

2. Visibility	
$\beta_{\text{vis}}^{4R|4R}$	-0.021 (-1.6)
$\beta_{\text{vis}}^{22L|22L}$	-0.027 (-2.12)
$\beta_{\text{vis}}^{29|29}$	-0.014 (-1.96)
$\beta_{\text{vis}}^{4R,11|4L}$	-0.013 (-2.67)
$\beta_{\text{vis}}^{4R,29|4L}$	-0.006 (-1.67)
$\beta_{\text{vis}}^{4L|4R}$	-0.025 (-2.73)
$\beta_{\text{vis}}^{22L,11|22R}$	-0.025 (-4.96)
$\beta_{\text{vis}}^{22L,29|22R}$	-0.03 (-2.2)
$\beta_{\text{vis}}^{22L,22R|22R}$	-0.009 (-3.1)
$\beta_{\text{vis}}^{22R,29|22L}$	-0.018 (-1.89)

3. Headwind speed	
$\beta_{\text{wind}}^{\text{prim,arr}}$	0.033 (3.0)
$\beta_{\text{wind}}^{\text{prim,dep}}$	0.054 (4.24)
$\beta_{\text{wind}}^{\text{ex,arr}}$	0.027 (3.93)

4. Demand	
$\beta_{\text{par}}^{\text{dem}}$	1.09 (8.33)
$\beta_{\text{ex\_arr}}^{\text{dem}}$	0.76 (5.9)
$\beta_{\text{ex\_dep}}^{\text{dem}}$	1.78 (7.15)

5. Noise abatement	
β_{11}^{mor}	-1.67 (-7.35)
β_{29}^{mor}	-1.86 (-6.6)

6. Switch type	
$\beta_{\text{switch}}^{1}$	-0.98 (-3.73)
$\beta_{\text{switch}}^{2}$	-0.62 (-1.8)
$\beta_{\text{switch}}^{3}$	-1.7 (-6.74)
$\beta_{\text{switch}}^{4}$	-2.23 (-3.23)
$\beta_{\text{switch}}^{5}$	-0.42 (-3.97)

7. Coordination with JFK	
$\beta_{4,31}^{\text{dep}}$	0.826 (2.35)
$\beta_{22,4}^{\text{dep}}$	-0.615 (-1.32)
$\beta_{22,13}^{\text{dep}}$	-1.14 (-2.35)
$\beta_{29,13}^{\text{dep}}$	-0.694 (-2.73)
$\beta_{4,22}^{\text{arr}}$	-1.25 (-3.07)
$\beta_{11,13}^{\text{arr}}$	0.437 (2.57)
$\beta_{11,31}^{\text{arr}}$	0.576 (2.84)
$\beta_{22,13}^{\text{arr}}$	1.2 (2.95)
$\beta_{22,22}^{\text{arr}}$	-0.94 (-2.63)
$\beta_{29,13}^{\text{arr}}$	1.13 (4.08)
$\beta_{29,22}^{\text{arr}}$	0.449 (1.66)
$\beta_{29,31}^{\text{arr}}$	1.22 (4.17)

8. Nest scale parameter	
$\mu_{\text{ex\_arr}}$	1.45 (3.23)

Table 3.4: Estimation results for configuration selection at EWR.

exhibit reasonable consistency across the two years, thereby corroborating their validity.

3.5.3.4 Illustration of implied configuration selection probabilities using hypothetical values for utility variables

For EWR, Figure 3-6 demonstrates the tradeoffs between switch proximity, demand-capacity inter-relationship and coordination with JFK in configuration selection, as implied by the parameter estimates. As with LGA, hypothetical scenarios are constructed controlling for other factors such as wind speed and direction (20 kn headwind for runway 11), visibility conditions (VFR), and noise abatement restrictions (not present). The focus is restricted to the prominent configurations (4R|4L; 4R,11|4L; 4R,29|4L; 22L|22R; 22L,11|22R; 22L|22R,29), and 4R|4L is assumed to be the incumbent configuration. Three demand scenarios are

considered (when demand exceeds crossing runway configuration capacity, when demand exceeds unhindered (or free) arrival capacity, and when demand exceeds parallel runway configuration capacity), each in conjunction with the two most prominent JFK runway configurations: 31R|31L and 13L|13R. As with LGA, the relative selection probabilities for the non-incumbent configurations are measured conditioned on a change in configuration.
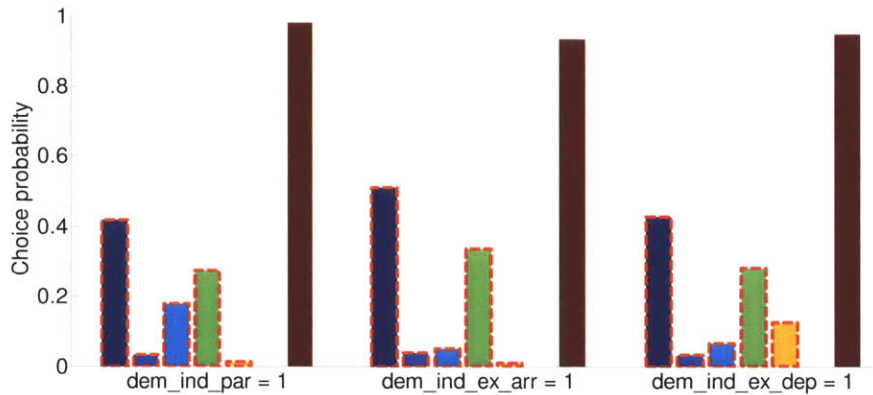
(a) JFK configuration:31R|31L
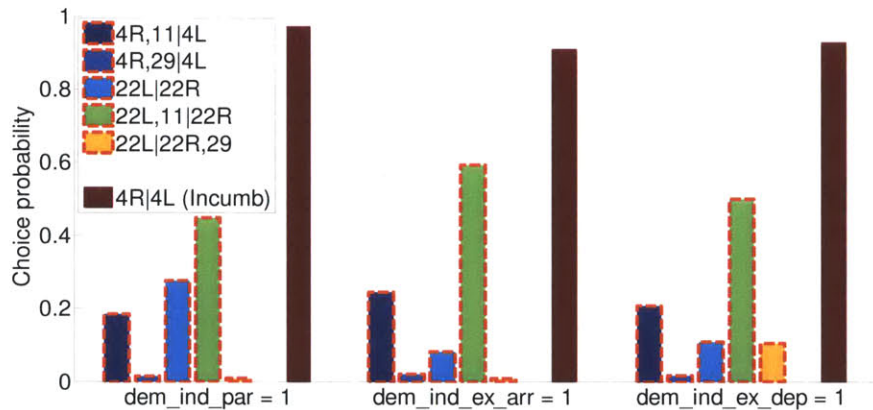
(b) JFK configuration:13L|13R

Figure 3-6: Relative configuration selection probabilities at EWR for described hypothetical scenarios.

When JFK is operating 31R|31L (Figure 3-6 (a)), it is noted that configuration 4R,11|4L dominates among the non-incumbents across all demand scenarios. This is due to its switching proximity relative to the incumbent, as well as favorableness given the JFK configura-

tion (note the positive values of coefficients for $\beta^{\text{arr}}_{11,31}$ and $\beta^{\text{dep}}_{4,31}$. Configuration 22L,11|22R is second-best due to the low utility associated with switch category 5: (180,180). Configurations featuring runway 29 are least preferred due to the adverse wind direction. Configurations with an additional arrival runway (like 4R,11|4L) are more preferable when demand exceeds unhindered arrival capacity (dem_ind$^{t}_{\text{ex\_arr}}$=1), while configurations with additional departure runway (like 22L|22R,29) are more preferable when demand exceeds parallel runway configuration capacity (dem_ind$^{t}_{\text{ex\_dep}}$=1). When JFK operates in 13L|13R (Figure 3-6 (b)), the dominant non-incumbent is 22L,11|22R, which is now favored by the JFK configuration (note that the coefficients $\beta^{\text{arr}}_{22,13}$ and $\beta^{\text{arr}}_{11,13}$ are both positive), overriding switch proximity considerations. Also, the preference for the configuration with additional departure runway (22L|22R,29) remains suppressed even when demand exceeds parallel runway configuration capacity, since the JFK configuration strongly inhibits it (negative signs of coefficients $\beta^{\text{dep}}_{29,13}$).

3.5.4 Model validation

This section describes the validation of the proposed discrete choice model for configuration selection and its parameter estimates. The validation analysis uses a test data set to compare the quality of configuration selection predictions between the estimated discrete choice model and a simpler model (termed the baseline model). The test set consisted of ASPM data records from 2007 for the study airports, refined using same filters applied for the training data set (2006 ASPM records). The baseline model structure is described in the next section, followed by a brief discussion of the validation results.

3.5.4.1 Baseline model

The use of the discrete choice modeling framework enables the incorporation of relevant influencing attributes like weather conditions, demand, etc. in determination of configuration selection probability. A simpler approach is to compute explicitly, using empirical observations, the probability of a particular configuration being chosen conditioned on the configuration used in the previous time interval. Such an approach effectively generates a

transition probability matrix Δ, where an element $\Delta_{i,j}$ represents the estimated probability of configuration j being chosen for any time interval t, given that configuration i was active in time interval $t-1$. Peterson (1992) [61] describes such a model of airport capacity dynamics based on a Markovian premise and featuring a finite number of capacity states, where each state represents a specific configuration. His empirical estimation procedure is used here to develop parameter estimates for the baseline model.

Given $C_t \ \forall t = \{1, 2, ...T\}; C_t \in \{1, ..., N_c\}$, where T is the total number of time intervals, N_c is the total number of possible configurations, and C_t is the selected configuration at time t,

$$\Delta_{i,j} = \frac{\sum_{t=1}^{T} (C_t == j) \wedge (C_{t-1} == i)}{\sum_{t=1}^{T} C_{t-1} == i} \quad \forall i, j \in \{1, .., N_c\}. \tag{3.11}$$

It can be shown that the above estimation framework is equivalent to a MNL discrete choice model where the configuration utilities are defined as the summation of $N_c - 1$ time-invariant categorical variables as expressed below.

$$V_{it} = \sum_{j \in \{1, .., N_c\}, j \neq i} \beta_{i,j}(\text{choice}_{t-1} = j) \tag{3.12}$$

Each categorical variable serves as an indicator of a specific runway configuration in the previous time-step, and the corresponding coefficient is specific to each configuration's utility. The key difference from the discrete-choice model proposed in this study is that other explanatory factors like weather, demand, etc. are not considered in the baseline model.

3.5.4.2 Baseline model estimates for LGA and EWR

Tables 3.5 and 3.6 present the estimated transition probability matrices (Δ) for LGA and EWR using the same dataset (2006 ASPM) as the discrete choice models. The tables only consider the most prominent configurations in both airports.

3.5.4.3 Validation results

In this study, prediction accuracy is assessed using aggregated configuration probabilities over the validation dataset. Since typical airport configuration planning horizons are of the order of 3 hours, the predicted probabilities are computed conditioned on the configura-

$\Delta_{i,j}$	4|4	31|31	4|13	22|13	22|31	31|4
4|4	0.941	0	0.021	0.011	0.002	0.018
31|31	0.001	0.957	0	0.002	0.019	0.019
4|13	0.004	0	0.974	0.013	0	0.004
22|13	0.001	0	0.005	0.980	0.007	0.003
22|31	0	0.005	0.001	0.008	0.972	0.009
31|4	0.003	0.003	0.007	0.003	0.008	0.977

Table 3.5: Baseline model estimates for LGA.

$\Delta_{i,j}$	4R|4L	4R,11|4L	4R,29|4L	22L|22R	22L,11|22R	22L|22R,29
4R|4L	0.983	0.006	0.003	0.006	0.001	0
4R,11|4L	0.019	0.972	0	0.005	0.003	0
4R,29|4L	0.019	0	0.973	0.007	0	0
22L|22R	0.004	0	0.001	0.987	0.003	0.004
22L,11|22R	0.004	0	0	0.016	0.976	0.003
22L|22R,29	0.001	0	0.001	0.040	0.013	0.941

Table 3.6: Baseline model estimates for EWR.

tion observed 3 hours before, and not the previous 15-min time period as modeled in the estimation process.

Suppose obs_conf$_t$ denotes the observed configuration for time-step t. The aggregate predicted probability (agg_pr$_i$) for configuration i using a 3-hr (that is, twelve 15-min periods) look-ahead is calculated as:

$$\text{agg\_pr}_i = \frac{\sum_{t:\text{obs\_conf}_t=i} P(\text{conf}_t = i | \text{conf}_{t-12} = \text{obs\_conf}_{t-12})}{\sum_{t:\text{obs\_conf}_t=i} 1}. \tag{3.13}$$

In (3.13), the 3-hr look-ahead prediction probability $P(\text{conf}_t = i | \text{conf}_{t-12} = k)$ is computed recursively in the following manner:

$$P(\text{conf}_t = j | \text{conf}_{t-12} = k) = \sum_{i=1}^{N_c} P(\text{conf}_t = j | \text{conf}_{t-1} = i) * P(\text{conf}_{t-1} = i | \text{conf}_{t-12} = k)$$

$$P(\text{conf}_{t-1} = i | \text{conf}_{t-12} = k) = \sum_{m=1}^{N_c} P(\text{conf}_{t-1} = i | \text{conf}_{t-2} = m) * P(\text{conf}_{t-2} = m | \text{conf}_{t-12} = k)$$

and so on.

The absolute prediction quality would naturally deteriorate as we increase the length of the look-ahead duration. However, it should not influence the relative comparison of

the models' prediction qualities. The validation results are presented below (Table 3.7 for LGA, and Table 3.8 for EWR). The results are partitioned for two disjoint data segments, the first comprising of observations from time periods that are not within 3 hours of the nearest observed configuration switch, and the second set comprising of the remainder of the dataset (i.e., within 3 hours before or after a switch). The results are presented for the most frequently used configurations at each airport, as listed in Tables 3.5 and 3.6. The validation tables show the aggregate probability of a runway configuration being correctly predicted, both near and away from configuration switches. The aggregate probabilities in the vicinity of a switch are conditioned on the event of a switch occurring. We note that a perfect prediction mechanism would deliver an aggregate probability equal to 1.

	Outside temporal vicinity of switches		
		Correct prediction	
Configuration	Frequency	Baseline	Discrete-Choice
22|13	4403	0.81	0.95
22|31	3725	0.73	0.92
31| 4	2989	0.77	0.90
4|13	2339	0.74	0.91
31|31	1211	0.61	0.70
4|4	599	0.50	0.69
	Within temporal vicinity of switches		
Configuration	Frequency	Baseline	Discrete-Choice
31|4	1103	0.48	0.71
22|31	1043	0.50	0.74
22|13	1024	0.55	0.76
4|13	569	0.47	0.58
31|31	403	0.31	0.57
4|4	135	0.31	0.44

Table 3.7: Validation results for LGA (aggregate probabilities of correct configuration prediction for 2007 dataset). Number of parameters in baseline model = 100; number of parameters in discrete-choice model = 36.

The validation results show that the predictions generated by the discrete-choice model are significantly better than those of the baseline model, in spite of the considerably smaller number of parameters required by the discrete-choice model. This result highlights the richer use of empirical information achieved by the discrete choice model. The fact that the improvement in prediction accuracy is consistent across the two disjoint sets of observations

| | Outside temporal vicinity of switches | | |
| | | Correct prediction | |
Configuration	Frequency	Baseline	Discrete-Choice
22L|22R	6583	0.88	0.87
4R|4L	4173	0.84	0.87
22L,11|22R	1686	0.77	0.94
4R,11|4L	1087	0.74	0.88
4R,29|4L	715	0.74	0.81
22L|22R,29	211	0.52	0.16
	Within temporal vicinity of switches		
Configuration	Frequency	Baseline	Discrete-Choice
22L|22R	2073	0.70	0.73
4R|4L	1303	0.65	0.73
22L,11|22R	799	0.32	0.76
22L|22R,29	573	0.24	0.21
4R,11|4L	505	0.40	0.74
4R,29|4L	336	0.29	0.70

Table 3.8: Validation results for EWR (aggregate probabilities of correct configuration prediction for 2007 dataset). Number of parameters in baseline model = 400; number of parameters in discrete-choice model = 57.

(near and away from configuration switches) demonstrates the superiority of the discrete-choice model in predicting both the timing of configuration switch as well as the retention of the incumbent configuration if the prevailing conditions don't motivate a switch. In general, the quality of prediction is lower in the vicinity of configuration switches due to the inertia term biasing predictions towards incumbent configurations. Similarly, it is noted that the model performs relatively poorly in predicting configurations that are used more infrequently.

3.6   Impact of Configuration Changes on Capacity

As seen in Section 3.2, airport authorities routinely revise active configurations guided by considerations such as wind, operational demand, noise mitigation, etc. Every configuration switch requires a reallocation of resources, as well as significant coordination and conformance monitoring. This effort can vary in magnitude and duration depending upon the nature of the switch and can interfere with operational efficiency, as briefly discussed earlier in this chapter. For instance, a configuration switch involving the addition of an extra arrival

or departure runway to boost current capacity would conceivably cause less disruption than a switch involving a complete turnaround of runway directions. The decrease in efficiency can be particularly detrimental when the switch coincides with a high-demand period. Past studies involving configuration planning have acknowledged the adverse impact of switches, and have implicitly accounted for it by assuming zero capacity for an arbitrarily fixed duration representing the switch, regardless of its type [27, 35, 53]. However, the duration of this impact has not previously been estimated in practice. The estimated impact of a configuration switch may be particularly inaccurate during high-demand periods, during which the errors in the underlying assumptions can get amplified.

This thesis proposes a systematic approach for estimating operational impacts of configuration switches that can improve understanding of airport operations and thereby the quality of capacity predictions. The objective is to derive estimates on both duration as well as capacity impacts of switches using ASPM data on airport throughput observations. The methodology explicitly distinguishes between switch types, as classified in Section 3.5.3.1, based on angular reorientation of respective runways. The approach is applied to three airports: EWR, JFK and DFW, using ASPM 15-min records for the year 2006, and the results and their implications for the operational characteristics of these airports are discussed.

3.6.1 Estimation methodology

The operational impact of configuration switch type k is described by two coupled attributes:

sw_dur_k : Duration for which switch effect lasts, and

sw_red_k : Percentage reduction in airport capacity during switch

Empirical evidence of these attributes will be found in the temporal vicinity of recorded configuration switches. The principal objective of the estimation process is to differentiate observations affected by switches from regular ones. The following considerations are important for the design of the estimation framework presented here.

Coupling of sw_red_k and sw_dur_k : In this study, the observed airport throughput serves as the common indicator for both switch duration and the switch-induced capacity reduction. The empirical manifestations of these two attributes are therefore coupled, and estimation methods would require fixing the values for one of them. A two-stage

framework is developed in this study. The first stage looks to detect the presence of operational inefficiency (sw_red$_k$) in the neighborhood of switches, while fixing the switch duration (sw_dur$_k$) to a reasonable constant. The second stage focuses on those switch types that exhibit statistically significant inefficiencies in the first stage, and replicates the first stage estimation procedure to compute capacity reduction magnitude for different values of switch duration, thereby generating a comprehensive measure of switch impact.

Potential reporting errors: ASPM data on airport operations is subject to reporting errors concerning the exact time of configuration switches. This issue implies that observations in the vicinity of recorded switches need not always provide a reliable indicator of associated operational inefficiencies. Hence, in addition to a localized analysis for switch effects, a global estimation approach that does not rely upon the actual reported switch instants is also used in first stage of the estimation framework. The local and global estimates are together used to derive a robust inference on the presence of switch-induced inefficiencies.

In this study, airport operational efficiency is assessed using arrival throughputs alone, given that it is the prioritized mode of operation over departures. The quantities used in the discussion, along with their definitions, are presented first.

$$\text{sw}_{n,t}^k : \quad \begin{cases} 1, & \text{if switch type } k \text{ recorded at time } t \text{ for day } n \text{ in dataset,} \\ 0, & \text{otherwise.} \end{cases}$$

where $t \in \{1, .., T_n\}$ (total number of time steps in n^{th} day),

$n \in \{1, .., N\}$ (total number of days in dataset),

$k \in \{1, .., K\}$ (total number of switch types for the given airport)

$\text{op}_{n,t}$:   Observed arrival throughput for time t for day n

$\text{dem}_{n,t}$:   Recorded arrival demand for time t for day n

$\text{cap}_{n,t}$:   Estimated arrival capacity (from Chapter 2) for time t for day n

$\text{target}_{n,t}$: $\min(\text{dem}_{n,t}, \text{cap}_{n,t})$, Operational target for airport for time t on day n

$\text{op\_eff}_{n,t}$: $\quad \frac{\text{op}_{n,t}}{\text{target}_{n,t}}$, Measure of operational efficiency achieved by airport at time t on day n

$\text{target\_perc}_{n,t}$: $\quad \frac{\text{target}_{n,t}}{\text{cap}_{n,t}}$, Measure of operational load faced by airport at time t on day n

Γ^k: $\quad \bigcup\limits_{n,t \,|\, \text{sw}_{n,t}=1} \{nt - \frac{\text{sw\_dur}_k}{2}, nt - \frac{\text{sw\_dur}_k}{2} + 1, ..., nt + \frac{\text{sw\_dur}_k}{2}\}$, set of time instances in local neighborhood of all recorded instances of switch type k assuming fixed switch duration sw_dur_k.

Λ: $\quad \bigcup\limits_{n,t} \{nt\}$, set of all recorded instances

Given the definition of the quantities, the estimation framework looks to quantify the trends in $\text{op\_eff}_{n,t}$ with respect to the influencing entity $\text{sw}_{n,t}^k$, $\forall k$.

3.6.1.1 First stage

As mentioned earlier, the first stage of the estimation fixes sw_dur_k, and uses local and global trends in $\text{op\_eff}_{n,t}$ with respect to $\text{sw}_{n,t}^k$ to ascertain the presence of tangible impacts of switch type k on operational efficiency.

3.6.1.1.1 Local Estimation: The local estimation attempts to statistically compare the mean operational efficiency $\text{op\_eff}_{n,t}$ observed within the switch neighborhood (Γ^k) to that observed outside it. This can be achieved by running a simple linear regression model of the following form.

$$\text{op\_eff}_{n,t} = \alpha_{\text{local}} + \sum_{k \in \{1,..,K\}} \beta_{\text{local}}^k \text{ind}_{n,t}^k + \epsilon_{n,t} \qquad \forall n, t \in \Lambda \qquad (3.14)$$

where indicator variable $\text{ind}_{n,t}^k = \{1, \text{ if } n, t \in \Gamma^k; \quad 0, \text{ otherwise}\}$. Here, the estimate for regression parameter β_{local}^k gives a measure of reduction in mean operational efficiency within the neighborhood of switch type k in the dataset, and its statistical significance is easily verified using its t-statistic.

3.6.1.1.2 Global estimation The global estimation attempts to obtain a similar measure of switch-induced mean operational inefficiency, but without a priori identification of switch neighborhood, thereby rendering the estimate impervious to potential inaccuracies in reported timing of switch instances. This is achieved through the following day-based linear regression model.

$$Y_n = \alpha_{\text{global}} + \sum_{k \in \{1,..,K\}} \beta^k_{\text{global}} X^k_n + \epsilon_n \qquad \forall n \in \{1,..,N\} \tag{3.15}$$

Here, the dependent variable $Y_n = \dfrac{\sum_{t \in \{1,..,T_n\}} \text{op\_eff}_{n,t}}{T_n}$ (average observed operational efficiency for the nth day), and the independent variable $X^k_n = \sum_{t \in \{1,..,T_n\}} \text{sw}^k_{n,t}$ (the total number of switches of type k recorded on the nth day). The regression parameter β^k_{global} measures switch impacts on an aggregated scale, and can also be verified for its statistical significance.

3.6.1.1.3 Inference rule: The following inference matrix (Table 3.10) describes how the local and global estimates of operational inefficiency, measured for fixed switch duration sw_dur_k, are jointly utilized to draw conclusions on the nature of operational impacts for each switch type k. Consistently significant measures from local and global analyses is inferred

	β^k_{global} is significantly negative	β^k_{global} is not significantly negative
β^k_{local} is significantly negative	Confirmed presence of switch impacts; warrants second stage estimation	Implies insignificance of switch impacts on daily aggregates; little value in further investigation
β^k_{local} is not significantly negative	Suggests possible errors in reporting of switch times; confounds further investigation	Confirmed absence of switch effect; no need for further investigation

Table 3.10: Inference matrix for first-stage analysis on configuration switch effects.

as providing conclusive evidence for the presence of switch effects, which are then further analyzed in the second stage of estimation.

3.6.1.2 Second stage

For a switch type k deemed to impose significant operational inefficiencies in the first stage analysis, the second stage looks to explicitly estimate the parameter sw_red_k for different values of sw_dur_k by executing a detailed version of the day-based regression developed for the global analysis in the first stage. While the local and global regression models in the first stage provide approximate measures of switch-induced mean operational inefficiency, the second stage focuses on estimating the percentage reduction in airport capacity per unit time during switches. The latter is a more instructive metric that can be readily used in models for configuration planning [27, 53]. The following linear regression model is adopted for the second stage estimation of sw_red_k for each assumed value of sw_dur_k.

$$Y_n = \alpha_{\text{II}} + \sum_{k \in \{1,..,K\}} \beta_{\text{II}}^k \frac{\text{sw\_dur}^k X_n^k}{T_n} + \epsilon_n \qquad \forall n \in \{1,..,N\} \tag{3.16}$$

where variables Y_n and X_n^k are as defined in 3.15 for global regression model from first stage. Given the choice of the independent variable $\frac{\text{sw\_dur}^k X_n^k}{T_n}$, which captures the fraction of the nth day's observations affected by switches, the regression coefficient β_{II}^k provides a direct measure of switch parameter sw_red_k.

The above regression model is used to enumerate the estimates for sw_red_k for different fixed values of sw_dur_k, to obtain insight into the measured switch effects. It is noted here that the regression models in (3.15) and (3.16) define dependent and independent variables based on daily averages of varying sample sizes T_n, $\forall n \in \{1,..,N\}$. Given this feature, weighted least-squares estimation (WLS) is adopted to restore homoskedasticity amongst the error terms, ϵ_n, and ensure efficient estimates for regression coefficients, as detailed in [62].

3.6.1.2.1 Controlling for operational loading parameter, $\text{target\_perc}_{n,t}$: Finally,

it is noted that the estimated measures for switch effects from the first and second stages can be further refined by controlling for parameter $\text{target\_perc}_{n,t}$ in the estimation dataset. The reasoning here is that switch effects are expected to manifest more visibly during times of high load, that is, values of $\text{target\_perc}_{n,t}$ close to 1. The estimation datasets for all the

above-discussed regression models (3.14), (3.15) and (3.16) are thus filtered based on varying thresholds of the parameter $target\_perc_{n,t}$, and focus is mainly devoted to the higher ranges of this parameter during each of the respective analysis in the two estimation stages.

3.6.2 Case studies: EWR, JFK and DFW

The two-stage estimation framework was applied to study configuration switch impacts at EWR, JFK and DFW using ASPM 15-min records from year 2006. Along with reported data on airport configuration, arrival throughput as well as demand for every 15-min observation, the unhindered arrival capacity estimate obtained using the quantile regression method from Chapter 2 is used for the analysis. Overnight hours are excluded from the estimation dataset as they are not representative of regular airport operations.

While the runway layout for EWR was depicted in Figure 3-5, the equivalent sketches for JFK and DFW are provided in Figure 3-7 and 3-8 respectively. There were a total of 20 and 24 distinct configurations observed at EWR and JFK respectively, while two distinct configurations involving opposite runway orientations (along the 18s and along the 36s) appeared in 85% of observations at DFW.

The switch categories defined in Section 3.5.3.1 of the configuration selection model, based on the angle of runway reorientation, are invoked for EWR and JFK in this analysis. Due to limited instances of individual categories in the dataset, switch classes 1 - 4 (i.e., all switches not involving a complete airport turn-around) are consolidated and termed the "base" switch class. The "non-base" switch class is one that requires a 180 eorientation of both arrival and departure runways. For DFW, this is the only form of configuration switch analyzed due to the predominance of two configurations in the dataset.

3.6.2.1 First stage analysis

The following tables display the local and global analysis results for the EWR (table 3.11), JFK (table 3.12) and DFW (table 3.13) airports respectively. Each table provides the respective estimates for switch effects (and the t-statistic in paranthesis) for all analyzed switch

Figure 3-7: JFK runway layout

Figure 3-8: DFW runway layout

types across different ranges of the operational loading parameter $target\_perc_{n,t}$. The switch duration sw_dur_k is set at five 15-min time periods (i.e. 1.25 hours) for all switch types assessed in the first stage.

$target\_perc_{n,t} \geq$	0.1	0.4	0.7	1.0
α_{local}	84.1 (724.23)	82.2 (666.1)	76.7 (549.2)	70.6 (478.1)
$\beta_{\text{local}}^{\text{base}}$	-0.34 (-0.67)	-1.23 (-2.15)	-2.81 (-4.21)	-7.34 (-9.51)
$\beta_{\text{local}}^{\text{non-base}}$	2.86 (8.17)	3.15 (8.31)	3.13 (6.98)	-0.52 (-0.94)
α_{global}	83.9 (102.7)	82.5 (99.8)	78.9 (99.22)	75.55 (126.5)
$\beta_{\text{global}}^{\text{base}}$	-3.72 (-5.97)	-3.99 (-5.92)	-4.49 (-6.19)	-5.27 (-6.46)
$\beta_{\text{global}}^{\text{non-base}}$	1.97 (3.96)	2.02 (3.89)	1.86 (3.35)	0.36 (0.55)
No. of base switches	199	189	169	93
No. of non-base switches	454	441	385	195

Table 3.11: First stage estimation results for EWR using 2006 ASPM data.

$target\_perc_{n,t} \geq$	0.1	0.4	0.7	1.0
α_{local}	91.4 (796.7)	85.8 (469.4)	70.6 (199.3)	59.7 (136)
$\beta_{\text{local}}^{\text{base}}$	-2.58 (-9.31)	0.11 (0.29)	4.48 (5.99)	2.18 (2.1)
$\beta_{\text{local}}^{\text{non-base}}$	-0.51 (-1.35)	1.4 (2.44)	7.82 (7.38)	6.63 (4.40)
α_{global}	86 (61.6)	86.0 (54.4)	89.1 (69.7)	80.1 (57.12)
$\beta_{\text{global}}^{\text{base}}$	1.66 (3.00)	0.59 (0.87)	-5.33 (-6.13)	-7.34 (-4.98)
$\beta_{\text{global}}^{\text{non-base}}$	2.2 (3.68)	1.30 (1.62)	-3.71 (-3.00)	-5.89 (-2.91)
No. of base switches	654	602	347	136
No. of non-base switches	324	275	146	53

Table 3.12: First stage estimation results for JFK using 2006 ASPM data.

$target\_perc_{n,t} \geq$	0.1	0.4	0.7	1.0
α_{local}	98.2 (1898.7)	98.4 (1774.1)	97.1 (675.8)	81.1 (64.6)
$\beta_{\text{local}}^{\text{non-base}}$	-4.26 (-15.9)	-6.39 (-19.55)	-9.64 (-12.9)	-9.73 (-2.43)
α_{global}	98.2 (360.9)	98.3 (331.2)	97.6 (229.4)	95.8 (100.7)
$\beta_{\text{global}}^{\text{non-base}}$	-0.97 (-2.68)	-1.49 (-3.24)	-2.15 (-2.70)	-12.65 (-4.38)
No. of non-base switches	135	101	73	18

Table 3.13: First stage estimation results for DFW using 2006 ASPM data.

It is noted that the switch types that register consistent measures of operational inefficiency on the local and global assessments also demonstrate a marked increase in the estimated inefficiency for higher ranges of $target\_perc_{n,t}$. This is worth noting, as the absence of such a trend with respect to $target\_perc_{n,t}$ could be another indication for the lack of operational impacts for a given switch type.

92

The corresponding inferences derived from the first-stage estimation results, based on the guidelines described in Section 3.6.1.1.3, are summarized in Tables 3.14, 3.15 and 3.16.

Switch type (k)	β^k_{local}	β^k_{global}	Inference
Base	Significantly negative	Significantly negative	Warrants second stage estimation
Non-base	Not significantly negative	Not significantly negative	Confirmed absence of impact

Table 3.14: Inferences from first stage results for EWR.

Switch type (k)	β^k_{local}	β^k_{global}	Inference
Base	Not significantly negative	Significantly negative (for higher ranges of target_perc$_{n,t}$)	Possible errors in reported switch times, no further investigation
Non-base	Not significantly negative	Significantly negative (for higher ranges of target_perc$_{n,t}$)	Possible errors in reported switch times, no further investigation

Table 3.15: Inferences from first stage results for JFK.

Switch type (k)	β^k_{local}	β^k_{global}	Inference
Non-base	Significantly negative	Significantly negative	Warrants second stage estimation

Table 3.16: Inferences from first stage results for DFW.

As highlighted in the inference tables, the "base" switch type at EWR and the "non-base" switch type at DFW were the only ones that justified further exploration through second stage estimation. For both "base" as well as "non-base" switch types at JFK, the local estimates of switch effects conflicted with the global estimates, suggesting that there were considerable errors in the reported switch times. This shortcoming in the dataset would undermine the reliability of the second stage estimation.

For EWR, it appears counter-intuitive that the "base" switch type exhibits significant operational impact while the "non-base" switch type, which one would expect to be more disruptive, does not. A plausible explanation for this observation is that the EWR authorities try to avoid severe operational disruptions during peak demand periods by planning their

93

"non-base" switches during hours of low airport demand, thereby mitigating the realized impacts.

3.6.2.2 Second stage analysis

The results for the estimated values of the switch parameter sw_red_k for different values of sw_dur_k, as obtained from the regression model in Equation 3.16, are presented in the form of line graphs for EWR "base" (Figure 3-9) and DFW "non-base" (Figure 3-10) switch types. The estimates are plotted for different ranges of operational loading parameter $target\_perc_{n,t}$, to study the sensitivity to demand levels. The estimated values of the capacity reduction due to the switch are rounded off to a minimum value of -100 %, which corresponds to a complete loss of capacity.
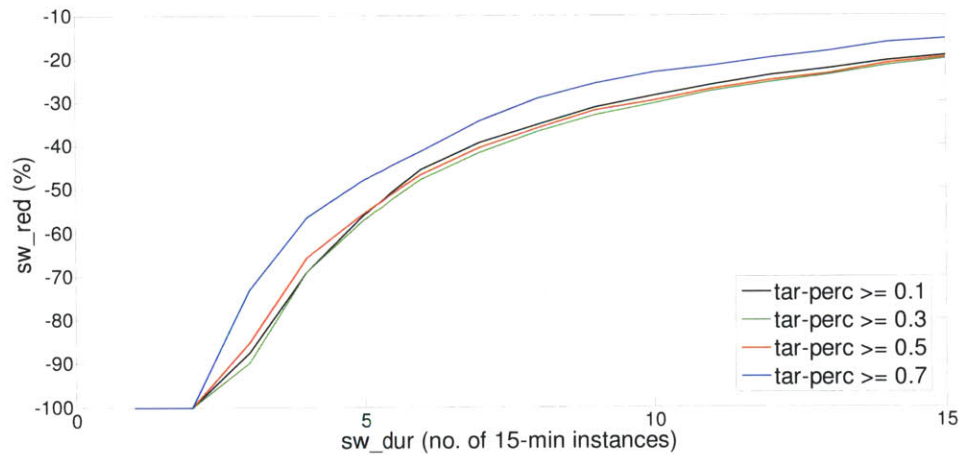
Figure 3-9: Second stage results for EWR "base" switches.

The plotted estimates of sw_red_k as a function of sw_dur_k are not very sensitive to variations in the value of $target\_perc_{n,t}$, especially in the case of DFW. This suggests that the extent of the capacity reduction does not depend strongly on the level of demand. The estimated percentage capacity reduction exhibits an inverse relation with assumed switch duration, which is a natural artifact of the coupling between these two attributes in the empirical observations. In other words, if one assumes a higher duration for switches, the estimated mean reduction in operational efficiency is likely to be lower when fitted to the
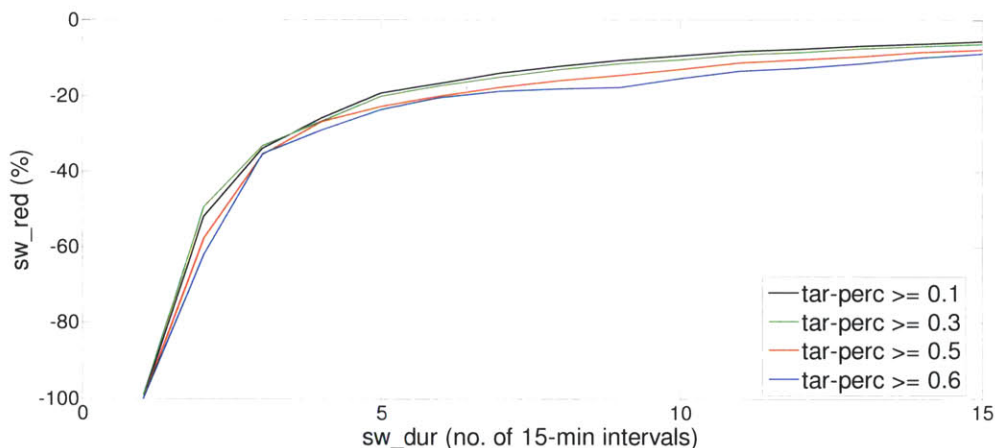
Figure 3-10: Second stage results for DFW "non-base" switches.

same empirical dataset.

3.6.2.3   Discussion of results

A couple of noteworthy deductions from the plots are highlighted below.

- The $x-$intercept of the line graphs gives an indication of the minimum duration of total capacity loss required for the respective switches. As shown in the figures, this value is relatively higher for the EWR "base" switches (between 15 and 30 min) compared to DFW "non-base" switches (within 15 min). This suggests a greater disruption due to a configuration change to operations at EWR, as compared to DFW.

  However, for EWR "base" switches, the t-stats for plotted estimates of parameter sw_red_k were between 3 and 4 across the range of values for $target\_perc_{n,t}$ and sw_dur_k, while the equivalent range for the DFW "non-base" switches was between 5 and 7.5. These values imply a lower statistical confidence in the switch impact estimates for EWR, potentially due to the presence of reporting errors and other noise in the ASPM data. This feature has to be remembered when assessing the plotted estimates for switch-induced capacity reduction sw_red_k for EWR "base" switches.

- The line graphs seem to taper out at values lower than 0% for high values of sw_dur_k,

95

suggesting that there is some residual, possibly spurious, operational inefficiencies within the reported ASPM throughput counts at both airports, that cannot be attributed to switch effects alone.

In summary, it is noted that any reliable pair of corresponding estimates for sw_red_k and sw_dur_k selected from the plotted line graphs will serve as a useful proxy for representing switch effects in applications involving configuration sequence modeling [27, 53]. In conjunction with the airport capacity estimation models from Chapter 2, these models of configuration dynamics help complete the modeling framework for airport operations that can now be used to generate realistic capacity forecasts. The following chapters address the problem of optimally allocating airport arrival and departure capacity under uncertainty.

Chapter 4

Integrated Stochastic Ground-Holding Problem

4.1 Introduction

Ground-holding is the practice of delaying a flight pre-departure at its origin airport to relieve congestion at its destination airport, thereby avoiding more expensive airborne delays. Airports initiate a ground delay program (GDP) when their capacity is expected to deteriorate in the near future, and assign delayed slots based on ground-holds for scheduled flights to ensure capacity is not exceeded.

As described in the introduction chapter, GDPs in practice are executed using deterministic capacity forecasts, and cater exclusively to arrivals only. Airports deal with the inherent uncertainty in future capacity by revising their slot allocation with every forecast update. Each GDP revision brings about a change in ground delays for scheduled slots, prompting a response from operating airlines through the CDM (Collaborative Decision-Making) mechanisms [18]. Airlines reconsider their flight schedules and perform slot swaps and cancellations to optimize their internal costs. Thus, a new iteration of intra-airline slot substitution and compression is executed for every GDP revision. This iterative framework is reactive in nature and contains inefficiencies that can be resolved by explicitly accounting for uncertainty. A stochastic model for ground-holding is based on the principle of representing capacity uncertainty through a discrete set of possible scenarios. The solution to this

model supplies airlines with advance information on the uncertainty in slot delays that they can use to proactively swap and cancel slots. Therefore, if reliable probabilistic capacity forecasts are available, stochastic models for ground-holding allocation can help deal with capacity uncertainty.

Another potential drawback of current GDP practice is the arrival-centric approach. Every GDP revision determines arrival slots over the declared duration, following which departures are handled in an ad-hoc manner. In other words, while arrival schedules are planned systematically during GDPs, departures are serviced based on the remaining airport capacity following arrival slot allocation. While such preferential treatment to arrivals during periods of capacity shortage is motivated by the desire to minimize airborne delays, it may cause excessive departure delays as a result of the capacity tradeoffs studied in Chapter 2.

This chapter describes integrated formulations for stochastic ground-holding models that address the above issues with existing GDP designs, and uses case studies involving real-world data to assess the potential benefits. The next section discusses related efforts that serve as the building blocks for this research.

4.2   Related Literature

The problem of optimally allocating ground delays to minimize system congestion costs has been extensively studied in literature [59, 74, 13]. While the single-airport ground-holding problem (SAGHP) has received the most attention, a few previous efforts [74, 13] have addressed aircraft flows over a network of airports, in which arrivals and departures from each airport are controlled in an integrated fashion. Bertsimas and Patterson (1998) used airport capacity envelopes to constrain simultaneous capacities for arrivals and departures [13]. In addition, Gilbo (1993) considered the problem of optimally allocating a single airport's capacity between arrival and departure demands [31]. However, these past studies on integrated capacity allocation all assumed deterministic capacity forecasts [31, 13].

A stochastic ground-holding approach is adopted when the magnitude or duration of capacity deterioration is not known with certainty at the start of a GDP. This problem has been formulated in literature as a stochastic integer program, where optimal ground-holds

are determined for future airport capacity that is predicted to materialize from a set of finite scenarios with associated probabilities [67, 4, 55, 36]. Most studies on stochastic ground-holding focus on single airports and typically assume that the airport operates at maximum arrival capacity at all time intervals. The underlying rationale is that the stochastic ground-holding problem, unlike its deterministic variant, yields airborne delays under some capacity scenarios. Since airborne delays are costlier than ground delays, these approaches reduce arrival (airborne) delays at the expense of departure (ground) delays. To the best of our knowledge, the simultaneous allocation of arrival and departure capacities within a stochastic framework has not been previously considered.

This chapter develops and assesses an integrated framework for stochastic ground-holding, thereby evaluating traditional arrival prioritization policy. This analysis is therefore a natural application for the airport capacity envelopes estimated in Chapter 2. Section 4.3 discusses capacity sharing in a deterministic setting to help explain why prioritizing arrivals might not be always be optimal under capacity uncertainty. Section 4.4 develops integrated versions of two prominent stochastic ground-holding models from literature, the *Static*, developed in Richetta and Odoni (1993) [67], and the *Dynamic*, proposed by Mukherjee and Hansen (2007) [55]. While the static model determines a single ground-holding solution for arrivals that is applicable across all capacity scenarios, the dynamic model revises ground holds at each time-step based on available information on capacity materialization. Section 4.4.3 describes how the use of non-dominated operating points to represent the airport capacity envelope in the integrated stochastic formulations enables customized, deeper branching cuts for the branch and bound solution algorithm. Section 4.6 presents case studies comparing the performances of the integrated and arrivals-based approaches, when applied to hypothetical and real-world GDP data. The influences of key GDP parameters such as demand magnitude and capacity forecasts are also investigated.

A key practical challenge in implementing the integrated stochastic models is the need to account for potential aircraft and passenger connections between arrivals and departures at the affected airport. Currently, GDPs are implemented within a CDM paradigm that provides mechanisms for airlines to participate in the slot allocation process and further reduce their delay costs in accordance to flight-specific preferences [18]. As information

on aircraft and passenger connections between scheduled flights is typically airline-specific, potential extensions could be designed within the CDM mechanisms that allow airlines to recover lost connections in the original integrated slot allocation. Related approaches have been proposed by Gilbo (2000) and Hall (1999) [30, 35].

4.3   Capacity Sharing with Deterministic Forecasts

Let us consider an idealized, single time-step example involving arrival demand A^d and departure demand D^d, at an airport with an operational capacity envelope as depicted in Figure 4-1. The capacity envelope is represented as a convex, piecewise-linear function [13, 31].
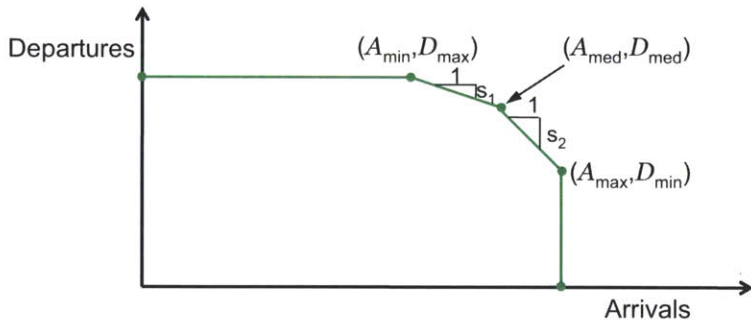
Figure 4-1: Capacity envelope for idealized example.

Let us denote the unit ground delay cost (for both arrivals and departures) and the unit airborne delay cost (only relevant for arrivals) as C_g and C_a respectively, with $C_a > C_g$. Let the linear pieces of the capacity envelope be such that:

$$s_1 < 1, \quad s_2 > 1, \quad A^d \geq A_{\max} \quad \text{and} \quad D^d > D_{\text{med}}. \tag{4.1}$$

Now, assuming that the arrivals are all airborne and awaiting landing clearance at the airport, the optimization formulation for minimizing system delay costs over a single time-

step is as described below:

$$\text{Minimize} \quad C_g D^{\text{gq}} + C_a A^{\text{aq}} \tag{4.2}$$

$$\text{subject to:} \quad D^{\text{gq}} \geq D^d - D^{\text{cap}} \tag{4.3}$$

$$A^{\text{aq}} \geq A^d - A^{\text{cap}} \tag{4.4}$$

$$A^{\text{cap}} \leq A_{\text{max}} \tag{4.5}$$

$$s_1 A^{\text{cap}} + D^{\text{cap}} \leq s_1 A_{\text{min}} + D_{\text{max}} \tag{4.6}$$

$$s_2 A^{\text{cap}} + D^{\text{cap}} \leq s_2 A_{\text{med}} + D_{\text{med}} \tag{4.7}$$

$$D^{\text{cap}} \leq D_{\text{max}} \tag{4.8}$$

$$A^{\text{aq}}, D^{\text{gq}}, A^{\text{cap}}, D^{\text{cap}} \geq 0 \tag{4.9}$$

where $A^{\text{cap}}, D^{\text{cap}}$ are the capacities allotted to arrivals and departures respectively, and $A^{\text{aq}}, D^{\text{gq}}$ are the unserved arrivals and departures at the end of the time-step, that enter the respective airborne and ground-held queues.

If $\frac{C_a}{C_g} > s_2$, it can be shown that the optimal solution for the above formulation will yield the capacity mix $A^{\text{cap}} = A_{\text{max}}$, $D^{\text{cap}} = D_{\text{min}}$. In other words, if the ratio of the airborne delay cost to ground delay cost is higher than the steepest slope of the convex capacity envelope, the optimal operating policy of the airport for a given time-step is to serve as many arrivals as possible.

We note, however, that the above result is only applicable to a single time-step situation in which the arrival demand cannot be ground-held. In a more realistic example where capacity forecast and GDP planning extends to multiple time steps $\{1, .., T\}$ into the future, arrival demand can be ground-held at origin airports in response to the optimal capacity mix determined for future time steps. Assuming a deterministic capacity forecast, and the same capacity and demand for each time-step $t \in \{1, .., T\}$ as before, the extended formulation

101

for minimizing system delay costs over multiple time-steps is as follows:

$$\text{Minimize} \quad \sum_{t=1}^{T} (C_g(A_t^{\text{gq}} + D_t^{\text{gq}}) + C_a A_t^{\text{aq}}) \tag{4.10}$$

$$\text{subject to:} \quad D_t^{\text{gq}} \geq D^d + D_{t-1}^{\text{gq}} - D_t^{\text{cap}}, \ \forall t \in \{1, .., T\} \tag{4.11}$$

$$A_t^{\text{gq}} \geq A^d + A_{t-1}^{\text{gq}} - A_t^{\text{arr}}, \ \forall t \in \{1, .., T\} \tag{4.12}$$

$$A_t^{\text{aq}} \geq A_t^{\text{arr}} + A_{t-1}^{\text{aq}} - A_t^{\text{cap}}, \ \forall t \in \{1, .., T\} \tag{4.13}$$

$$A_t^{\text{cap}} \leq A_{\text{max}} \tag{4.14}$$

$$s_1 A_t^{\text{cap}} + D_t^{\text{cap}} \leq s_1 A_{\text{min}} + D_{\text{max}} \tag{4.15}$$

$$s_2 A_t^{\text{cap}} + D_t^{\text{cap}} \leq s_2 A_{\text{med}} + D_{\text{med}}$$

$$D_t^{\text{cap}} \leq D_{\text{max}}, \ \forall t \in \{1, .., T\} \tag{4.16}$$

$$A_t^{\text{gq}}, A_t^{\text{aq}}, D_t^{\text{gq}}, A_t^{\text{arr}}, A_t^{\text{cap}}, D_t^{\text{cap}} \geq 0 \tag{4.17}$$

where, A_t^{arr} is the number of aircraft arriving at time t, A_t^{gq} is the number of aircraft scheduled to land at or before time t that are kept in the ground-held queue at their origin, A_t^{aq} is the number of arrival aircraft in the airborne queue at time t, and D_t^{gq} is the number of departures in ground-held queue at time t.

Assuming that arrivals scheduled over future time steps $\{1, .., T\}$ can all be subject to ground-holding and given condition (4.1), it can be shown that the optimal solution for the multiple time-step formulation will yield the capacity mix $A_t^{\text{cap}} = A_{\text{med}}$, $D_t^{\text{cap}} = D_{\text{med}}$ $\forall t \in \{1, .., T\}$, and contain no airborne queue (that is, $A_t^{\text{aq}} = 0$ $\forall t \in \{1, .., T\}$). The magnitude of the cost ratio $\frac{C_a}{C_g}$ is irrelevant in this example. In other words, in a deterministic setting spanning a future time horizon for which scheduled arrivals are yet to take-off, the optimal operating policy for the airport is to maintain maximum *total* number of ground operations. This operating point corresponds to $A^{\text{cap}} = A_{\text{med}}$, $D^{\text{cap}} = D_{\text{med}}$.

We now consider a setting in which the capacity forecast is uncertain over future time steps, and scheduled arrivals are all yet to takeoff at their origin. In this setting, arrival airborne delays can occur in some time-steps due to an unexpected deterioration in capacity. Neither of the operating policies from the previously discussed single- or multi-time-step examples can be consistently adopted in a stochastic framework for minimizing system delay

costs. The optimal capacity mix in a stochastic framework can only be derived contingent on demand and capacity forecast data, and will generate lower system delay costs compared to any approach relying upon a pre-determined policy for the capacity mix. It is, however, important to note that the availability of accurate capacity and demand forecasts during a GDP would be instrumental in realizing these delay benefits.

The above conceptual argument serves to motivate the study of the integrated stochastic ground-holding problem, and the magnitude of delay benefits generated by it. The research seeks to understand the dependence of these delay benefits on key influencing factors in a GDP pertaining to demand and capacity forecasts, and to do so through experimental case studies simulating typical GDPs. Hypothetical GDPs with high arrival demand and airborne delay costs are designed that are most likely to favor the arrival prioritization policy derived for the single time-step example. Such a setting would offer a validation of the hypothesized benefit of capacity sharing in an integrated stochastic framework. In the next section, the formulations for integrated versions of the static and the dynamic stochastic ground-holding problems are presented. The computational advantages of using non-dominated operating points to represent the capacity envelope in these formulations are also discussed.

4.4   Integrated Stochastic Ground-Holding Models

Stochastic ground-holding models developed in literature allocate arrival slots while explicitly planning for uncertainty in capacity forecasts [67, 55]. In these models, arrival capacity is predicted to materialize as one out of a finite number of scenarios, with corresponding probabilities. A scenario tree depicts the time-steps at which each capacity scenario becomes distinguishable from others, as shown in Figure 4-2. The minimized objective function is the expected sum of ground and airborne delay costs across all scenarios, computed by assuming homogeneous unit costs for ground and airborne delays across all flights. In Sections 4.4.1 and 4.4.2, we discuss the integrated versions of two stochastic ground-holding models: Static [67] and Dynamic [55].
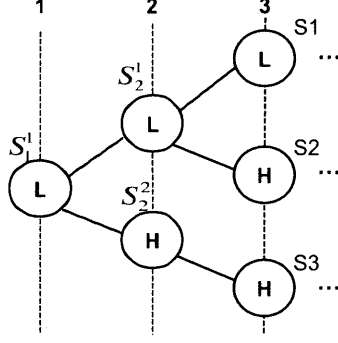
Figure 4-2: Sample capacity scenario tree.

4.4.1 Integrated Static model

As described in Richetta and Odoni (1993) [67], the arrivals-only static model is formulated as a single-stage stochastic integer program. The reader is referred to Wagner (1975) for a fundamental treatment of single-stage stochastic programs [39]. The first-stage decision variables are the ground delays allocated in aggregate form to arrivals scheduled within the GDP time horizon. In the static model, these decisions are fixed for the entire horizon at the start of the GDP, and are implemented regardless of the capacity scenario that eventually materializes.

The arrivals-only and the integrated versions of the static stochastic formulation are presented below, with the key additions in the integrated version highlighted.

Arrivals-only Static Formulation:

$$\text{Minimize} \quad (n-1)\sum_{n=0}^{K} C_g(\sum_{t=1}^{T-n} A_{t,t+n}^{\text{gq}}) + \sum_{s \in S} p_s(C_a \sum_{t=1}^{T} A_{s,t}^{\text{aq}}) \tag{4.18}$$

subject to:

$$\sum_{j=t}^{t+K} A_{t,j}^{\text{gq}} = A_t^d, \ \forall t \in \{1,..,T\} \tag{4.19}$$

$$A_{s,t}^{\text{aq}} \geq \sum_{j=t-K}^{t} A_{j,t}^{\text{gq}} + A_{s,t-1}^{\text{aq}} - A_{s,t}^{\text{cap}}, \ \forall t \in \{1,..,T\}, s \in S \tag{4.20}$$

$$A_{t,j}^{\text{gq}}, A_{s,t}^{\text{aq}} \in \mathbb{Z}^+, \ \forall t,j \in \{1,..,T\}, s \in S \tag{4.21}$$

104

Integrated Static Formulation:

$$\text{Minimize} \quad \sum_{n=0}^{K} C_{g,n} \sum_{t=1}^{T-n} A_{t,t+n}^{\text{gq}} + \sum_{s \in S} p_s \left[C_a \sum_{t=1}^{T} A_{s,t}^{\text{aq}} + \sum_{n=0}^{K} C_{g,n} \sum_{t=1}^{T-n} D_{t,t+n}^{\text{gq},s} \right] \quad (4.22)$$

$$\text{subject to:} \quad \sum_{j=t}^{t+K} A_{t,j}^{\text{gq}} = A_t^d, \ \forall t \in \{1,..,T\} \quad (4.23)$$

$$A_{s,t}^{\text{aq}} \geq \sum_{j=t-K}^{t} A_{j,t}^{\text{gq}} + A_{s,t-1}^{\text{aq}} - A_{s,t}^{\text{cap}}, \ \forall t \in \{1,..,T\}, s \in S \quad (4.24)$$

$$\sum_{j=t}^{t+K} D_{t,j}^{\text{gq},s} = D_t^d, \ \forall t \in \{1,..,T\}, s \in S \quad (4.25)$$

$$\sum_{j=\max(1,t-K)}^{t} D_{j,t}^{\text{gq},s} \leq D_{s,t}^{\text{cap}}, \ \forall t \in \{1,..,T\} \quad (4.26)$$

$$\alpha_{s,t}^{i} A_{s,t}^{\text{cap}} + \beta_{s,t}^{i} D_{s,t}^{\text{cap}} \leq \gamma_{s,t}^{i}, \ \forall i \in E_{s,t}, t \in \{1,..,T\}, \forall s \in S \quad (4.27)$$

$$A_{s^1,t}^{\text{cap}} = A_{s^2,t}^{\text{cap}}, \ \forall s^1, s^2 \in G(t,k), \forall t \in \{1,..,T\}, k \in \text{Stage}_t \quad (4.28)$$

$$D_{j,t}^{\text{gq},s^1} = D_{j,t}^{\text{gq},s^2}, \ \forall s^1, s^2 \in G(t,k), \forall j \in \{1,..,t\},$$
$$\forall t \in \{1,..,T\}, k \in \text{Stage}_t \quad (4.29)$$

$$A_{t,j}^{\text{gq}}, A_{s,t}^{\text{aq}}, D_{t,j}^{\text{gq},s}, A_{s,t}^{\text{cap}}, D_{s,t}^{\text{cap}} \in \mathbb{Z}^+, \ \forall t, j \in \{1,..,T\}, \forall s \in S \quad (4.30)$$

Notation

  Input

$C_{g,n}$:   Ground-delay cost incurred by an arrival or departure aircraft over n time-steps

K :   Maximum number of time-steps for which any aircraft can be ground-held.

C_a :   Linear unit airborne delay cost

A_t^d :   Aggregate arrival demand at time t

D_t^d :   Aggregate departure demand at t (used only in Integrated version)

Stage_t :   List of indices for distinct stages at time t in capacity scenario tree, with stage as described below (used only in Integrated version)

$G(t,k)$:   Subset of scenarios still possible at stage $k \in \text{Stage}_t$ (used only in Integrated version)

$E_{s,t}$	:	List of indices for linear pieces of capacity envelope under scenario s at time t (used only in Integrated version)
$\alpha_{s,t}^i, \beta_{s,t}^i, \gamma_{s,t}^i$	:	Coefficients of linear piece $i \in E_{s,t}$ of capacity envelope at scenario s and time t (used only in Integrated version)
p_s	:	Probability of occurrence for scenario s
S	:	List of possible capacity scenarios

Decision Variables

$A_{t,t+n}^{\text{gq}}$	:	Number of arrivals rescheduled from arrival time t to arrival time $t + n$ through ground-holding
$A_{s,t}^{\text{aq}}$	:	Length of arrival queue at time t for scenario s
$D_{t,t+n}^{\text{gq},s}$	:	Number of departures reschedule from departure time t to departure time $t + n$ under scenario s (used only in Integrated version)
$A_{s,t}^{\text{cap}}$	:	Airport capacity allotted to arrivals under scenario s at time t (set to maximum possible airport capacity for arrivals-only version)
$D_{s,t}^{\text{cap}}$	:	Airport capacity allotted to departures under scenario s at time t (used only in Integrated version)

Objective function

| (4.18) | : | Expected system delay cost for Arrivals-only version = Arrival ground delay cost + airborne delay cost |
| (4.22) | : | Expected system delay cost for Integrated version = Arrival ground delay cost + airborne delay cost + Departure ground delay cost |

Constraints

(4.19), (4.23) : Arrival demand balance

(4.20), (4.24) : Arrival capacity-queue balance

(4.25) : Departure demand balance (used only in Integrated version)

(4.26) : Depature capacity balance (used only in Integrated version)

(4.27) : Airport capacity envelope (used only in Integrated version)

(4.28) : Non-anticipativity constraint for allotted arrival capacity (used only in Integrated version)

(4.29) : Non-anticipativity constraint for allotted departure ground-holding (used only in Integrated version)

The integrated static stochastic model is formulated as a multi-stage stochastic integer program with the arrival ground delays ($A^{\text{gq}}_{t,t+n}$) as the first-stage decision variables, and with the capacity mix ($A^{\text{cap}}_{s,t}, D^{\text{cap}}_{s,t}$) and departure ground-delays ($D^{\text{gq},s}_{t,t+n}$) allocated in a stage-specific manner. In a capacity scenario tree, a stage is jointly defined by a time step t and the set of scenarios possible given past sequence of capacity values. A stage $k \in \text{Stage}_t$ uniquely identifies the set of possible future capacity scenarios that are indistinguishable based on the observed sequence of capacity values ($G(t,k)$). For instance, there are two distinct stages at time-step 2 for the example in Figure 4-2, with the stage S^1_2 comprising of $S1, S2$ and stage S^2_2 comprising of $S3$ as possible future scenarios, respectively. Non-anticipativity constraints (4.28) and (4.29) are enforced to ensure the stage-specific decisions on capacity mix and departure ground-delays are identical across all scenarios possible at a given stage. Among the other constraints are the standard demand and capacity balance constraints for arrivals and departures. Finally, (4.27) ensures the capacities allotted to arrivals ($A^{\text{cap}}_{s,t}$) and departures ($D^{\text{cap}}_{s,t}$) adhere to the available capacity envelope under scenario s at time t (that is, $E_{s,t}$).

Note that the constraints (4.29) require departures that are rescheduled from time j to time t to satisfy non-anticipativity for stages at time t instead of at time j, because the departure ground-hold decision $D^{\text{gq},s}_{j,t}$ is only determined at time t.

4.4.2 Integrated Dynamic model

In contrast to the static model, the dynamic model as developed in Mukherjee and Hansen (2007) [55] is a multi-stage stochastic mixed-integer program that determines stage-specific ground delay decisions for arrival flights that are still on the ground. The reader is referred to Kall (1976) for a comprehensive treatment of multi-stage stochastic programs [63]. The dynamic model thereby allows for scenario-specific determination of ground-holds, improving upon the optimal expected delay cost achieved by the static model. It also focuses separately on each individual flight, in contrast to the aggregate approach adopted in the static formulation, and makes explicit use of the flight's duration in allotting its scenario-specific ground-holds. The arrivals-only and integrated versions of the dynamic stochastic formulation are presented below.

Arrivals-only Dynamic Formulation:

$$\text{Minimize} \quad \sum_{s \in S} p_s \left[\sum_{f \in F} \left(\sum_{t=\text{arr}_f}^{\text{arr}_f + K} C_{g,t-\text{arr}_f} X_{f,t}^s \right) + \left(C_a \sum_{t=1}^{T} A_{s,t}^{\text{aq}} \right) \right] \tag{4.31}$$

$$\text{subject to:} \quad \sum_{t=\text{arr}_f}^{\text{arr}_f + K} X_{f,t}^s = 1, \ \forall s \in S, \forall f \in F \tag{4.32}$$

$$A_{s,t}^{\text{aq}} \geq \sum_{f \in F} X_{f,t}^s + A_{s,t-1}^{\text{aq}} - A_{s,t}^{\text{cap}}, \ \forall t \in \{1,..,T\}, s \in S \tag{4.33}$$

$$X_{f,t}^{s^1} = X_{f,t}^{s^2}, \ \forall s^1, s^2 \in G(t - \text{dur}_f) \tag{4.34}$$

$$X_{f,j}^s \in \{0,1\}, \ A_{s,t}^{\text{aq}} \in \mathbb{Z}^+, \ \forall t \in \{1,..,T\}, \ \forall s \in S, \ \forall f \in F \tag{4.35}$$

Integrated Dynamic Formulation:

$$\text{Minimize} \quad \sum_{s \in S} p_s [\sum_{f \in F} \sum_{t=\text{arr}_f}^{\text{arr}_f+K} C_{g,t-\text{arr}_f} X_{f,t}^s + C_a \sum_{t=1}^{T} A_{s,t}^{\text{aq}} + \sum_{n=0}^{K} C_{g,n}(\sum_{t=1}^{T-n} D_{t,t+n}^{\text{gq},s})] \quad (4.36)$$

$$\text{subject to:} \quad \sum_{t=\text{arr}_f}^{\text{arr}_f+K} X_{f,t}^s = 1, \; \forall s \in S, \; \forall f \in F \quad (4.37)$$

$$A_{s,t}^{\text{aq}} \geq \sum_{f \in F} X_{f,t}^s + A_{s,t-1}^{\text{aq}} - A_{s,t}^{\text{cap}}, \; \forall t \in \{1,..,T\}, \; s \in S \quad (4.38)$$

$$\sum_{j=t}^{t+K} D_{t,j}^{\text{gq},s} = D_t^d, \; \forall t \in \{1,..,T\}, \; s \in S \quad (4.39)$$

$$\sum_{j=\max(1,t-K)}^{t} D_{j,t}^{\text{gq},s} \leq D_{s,t}^{\text{cap}}, \; \forall t \in \{1,..,T\} \quad (4.40)$$

$$\alpha_{s,t}^i A_{s,t}^{\text{cap}} + \beta_{s,t}^i D_{s,t}^{\text{cap}} \leq \gamma_{s,t}^i, \; \forall i \in E_{s,t}, \; t \in \{1,..,T\}, \; s \in S \quad (4.41)$$

$$A_{s^1,t}^{\text{cap}} = A_{s^2,t}^{\text{cap}}, \; \forall s^1, s^2 \in G(t,k), \; \forall t \in \{1,..,T\}, \; k \in \text{State}_t \quad (4.42)$$

$$D_{j,t}^{\text{gq},s^1} = D_{j,t}^{\text{gq},s^2}, \; \forall s^1, s^2 \in G(t,k), \; \forall j \in \{1,..,t\}, \; t \in \{1,..,T\},$$
$$k \in \text{State}_t \quad (4.43)$$

$$X_{f,t}^{s^1} = X_{f,t}^{s^2}, \; \forall s^1, s^2 \in G(t-\text{dur}_f, k), \; \forall t \in \{1,..,T\},$$
$$\forall k \in \text{State}_{t-\text{dur}_f} \quad (4.44)$$

$$X_{f,j}^s \in \{0,1\}, A_{s,t}^{\text{aq}}, D_{t,j}^{\text{gq},s}, A_{s,t}^{\text{cap}}, D_{s,t}^{\text{cap}} \in \mathbb{Z}^+, \; \forall t \in \{1,..,T\},$$
$$\forall s \in S, \; \forall f \in F \quad (4.45)$$

Extended notation (in addition to that of the Static formulation)

Input

F : Set of flights scheduled to arrive at subject airport during the GDP horizon

arr_f : Originally scheduled arrival time interval for flight $f \in F$

dur_f : Duration for flight $f \in F$

Decision Variables

$X_{f,t}^s$: 1, if arrival flight f is rescheduled to arrive at time t for scenario s;
0, otherwise

Objective function

(4.31) : Expected system delay cost for Arrivals-only version = Arrival ground delay cost + arrival airborne delay cost

(4.36) : Expected system delay cost for Integrated version = Arrival ground delay cost + arrival airborne delay cost + departure ground delay cost

Constraints

(4.32), (4.37) : Arrival demand balance

(4.33), (4.38) : Arrival queue balance

(4.44) : Non-anticipativity constraints for arrival ground-holding

As shown above, the only difference between the formulations of the dynamic model and the static model is the structure for the arrival ground-holding decisions. Since the dynamic model determines stage-specific ground-hold for arrivals ($X_{f,t}^s$), an additional set of non-anticipativity constraints (4.44) is needed for these decision variables, along with that for capacity mix and departure ground-holds for the integrated version. Note that unlike the capacity mix ($A_{s,t}^{\mathrm{cap}}, D_{s,t}^{\mathrm{cap}}$) and departure ground-hold ($D_{t,t+n}^{\mathrm{gq,s}}$) decisions, the arrival ground-hold decision $X_{f,t}^s$ is determined at time $t - \mathrm{dur}_f$ and not t. This distinction is accordingly reflected in the non-anticipativity constraint (4.44), where the stages for $X_{f,t}^s$ are selected from $\mathrm{Stage}_{t-\mathrm{dur}_f}$ and not Stage_t.

The explicit use of the flight duration dur_f for ground-hold allocation in the dynamic model is also responsible for an inherent schedule rearrangement, as elaborated in Mukherjee and Hansen (2007) [55]. The final solution tends to favor long-haul arrivals over short-haul arrivals since the ground-hold decision on a lower duration flight is determined at a later time, when more specific information on capacity materialization is available. The shorter flights therefore receive a disproportionate share of the ground delayss under scenarios with prolonged low capacity. This schedule rearrangement poses equity concerns, as noted in Mukherjee and Hansen (2007) [55]. In the experimental case studies discussed later, the inequity in arrival ground-holds generated by the integrated dynamic model is compared

with the equivalent measure for the arrivals-only dynamic model.

4.4.3 Formulation properties

In recent work, Kotnyek and Richetta (2006) [48] proved the arrivals-only version of the static stochastic model ((4.18)-(4.21)) is guaranteed to have an integral optimum if the unit ground-hold delay costs $C_{g,n}$ are marginally non-decreasing (i.e., $C_{g,n+1} - C_{g,n} \geq C_{g,n} - C_{g,n-1}$, $\forall n$). The premise of this result was that the under the given cost conditions, at least one optimal arrival ground-hold solution exhibited a non-crossing structure (i.e., if for some $i, j \in \{1, .., T\}$ $A_{i,j}^{\text{gq}} > 0$, then $A_{l,k}^{\text{gq}} = 0$, $\forall i < l < k < j$).

For the given ground delay costs, it can be shown that the non-anticipativity constraints (4.29) for the departure ground-hold decisions $D_{j,t}^{\text{gq},s}$ are redundant in both the static and dynamic formulations.

An implication of the above result, combined with the original result from Kotnyek and Richetta (2006) [48], is that under marginally non-decreasing ground delay costs, the integrality requirement in the integrated static formulation can be limited to the capacity mix variables $A_{s,t}^{\text{cap}}, D_{s,t}^{\text{cap}}$. This simplification is possible because the determination of the capacity mix variables decouples the arrival and departure ground-holding sub-problems of the integrated model formulation. Using the above arguments, both sub-problems can be shown to yield integral optima under integral capacities and marginally non-decreasing ground delay costs. For the integrated dynamic formulation, we would additionally need to impose binary values on the arrival ground-holding decision $X_{f,t}^{s}$ as well.

4.5 Use of Non-Dominated Operating Points

In the formulations presented above for integrated versions of static and dynamic models, the capacity envelope was represented using a set of linear segments. Alternatively, it can be represented as the convex combination of its extreme points. Specifically, given the nature of the ground-holding formulations, such a representation can be restricted to the set of non-dominated points within the capacity envelope. A non-dominated operating point $A^{\text{cap}}, D^{\text{cap}}$ is defined such that there does not exist another feasible operating point a, d within the

capacity envelope with either $a > A^{\mathrm{cap}}, d \geq D^{\mathrm{cap}}$ or $a \geq A^{\mathrm{cap}}, d > D^{\mathrm{cap}}$. Figure 4-3 below illustrates the non-dominated operating points within a typical capacity envelope, estimated for operations over a 15 min interval. We note that the list of indices for non-dominated points, $i \in \mathbb{ND}$, can be arranged in increasing order of the underlying arrival (or departure) capacity $A^{\mathrm{cap},i}$ (or $D^{\mathrm{cap},i}$), $\forall i \in \mathbb{ND}$. For example, in Figure 4-3, there are five non-dominated integral operating points $\{A^{\mathrm{cap},i}, D^{\mathrm{cap},i}\} = \{(8, 14), (13, 13), (14, 10), (15, 8), (16, 3)\}$ indexed as $\mathbb{ND} = \{1, 2, 3, 4, 5\}$.
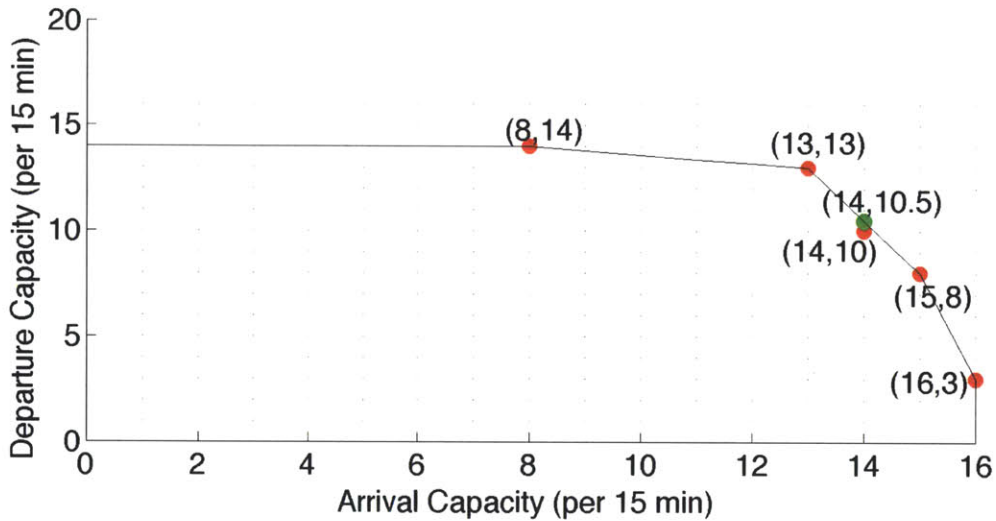
Figure 4-3: Example of a capacity envelope with the non-dominated operating points denoted by solid red circles.

For the stochastic ground-holding formulations discussed earlier, there exists at least one optimal solution that features a non-dominated operating point as the capacity mix solution $A_{s,t}^{\mathrm{cap}}, D_{s,t}^{\mathrm{cap}}$ for every scenario $s \in S$ and time $t \in \{1, .., T\}$. If there is an optimal solution with a dominated operating point at any stage of the scenario tree, replacing it with a non-dominated operating point that dominates it will not affect either the feasibility or the cost of the solution. Therefore, the capacity envelope can be represented as a convex combination of the set of non-dominated operating points $\mathbb{ND}$ for the integrated stochastic formulations,

112

as presented below.

$$A_{s,t}^{\text{cap}} \leq \sum_{i \in \mathbb{ND}_{s,t}} A_{s,t}^{\text{cap},i} \phi_{s,t}^i$$

$$D_{s,t}^{\text{cap}} \leq \sum_{i \in \mathbb{ND}_{s,t}} D_{s,t}^{\text{cap},i} \phi_{s,t}^i$$

$$\sum_{i \in \mathbb{ND}_{s,t}} \phi_{s,t}^i = 1, \ \forall t \in \{1, .., T\}, \ s \in S \tag{4.46}$$

The constraint set (4.46) replaces (4.27) and (4.41), respectively, in the integrated static and dynamic formulations. $\phi_{s,t}^i \in 0, 1, \ \forall i \in \mathbb{ND}_{s,t}$ are the binary coefficients of the non-dominated operating points in the convex combination.

The presented forms of the integrated static and dynamic models in Section 4.4 are both integer formulations, with the capacity mix variables $A_{s,t}^{\text{cap}}$, $D_{s,t}^{\text{cap}}$ being the primary source of integrality. For the integrated dynamic model, the arrival ground-hold decision variables $X_{f,t}^s$ also need explicit binary enforcement. For the alternative formulation using the non-dominated points, it is sufficient to enforce binary conditions for the variables $\phi_{s,t}^i, \ \forall i \in \mathbb{ND}_{s,t}$, to ensure integrality for the capacity mix variables, $A_{s,t}^{\text{cap}}, D_{s,t}^{\text{cap}}$.

Branch-and-bound is the most common solution algorithm used by commercial optimization packages for integer formulations [20]. Starting with the linear relaxation of the integer formulation, this algorithm employs the principle of iteratively sub-dividing the solution domain (known as branching), eliminating non-integral solutions and narrowing the search space, to help obtain integral optima for the sub-problems, which are then used to bound the optimal objective value of the original problem. The computational efficiency of the branch-and-bound algorithm is largely determined by the number of non-integral solutions eliminated in the branching step. Conventional branching schemes focus on a randomly chosen non-integer variable $X = X^* \notin \mathbb{Z}^+$ in the optimal solution for a given sub-problem, and enforce partitioning constraints $X \leq \lfloor X^* \rfloor$ and $X \geq \lceil X^* \rceil$ to further sub-divide the sub-problem.

The use of non-dominated operating points facilitates customized, stronger branching schemes. Consider, for example, the capacity envelope presented in Figure 4-3 and an optimal capacity mix of $A^{\text{cap},*} = 14$, $D^{\text{cap},*} = 10.5$ obtained by linear relaxation, for some given sce-

113

nario and time step. For the original piecewise-linear representation of the capacity envelope, conventional schemes would branch on the variable $D^{\mathrm{cap},*}$, and produce sub-problems through partitioning constraints $D^{\mathrm{cap}} \leq 10$ and $D^{\mathrm{cap}} \geq 11$ respectively. For the non-dominated points representation, adopting the above-described indexing for set $\mathbb{ND}$, the values for the relevant binary variables would be $\phi^2 = 0.5$, $\phi^4 = 0.5$ and $\phi^i = 0$, $\forall i \in \{1, 3, 5\}$. A potential branching scheme could use partitioning constraints $\sum_{i \in \{1,2\}} \phi^i = 1$, and $\sum_{i \in \{3,4,5\}} \phi^i = 1$ for generating the sub-problems. Note that this branching scheme makes explicit use of the ordering of the non-dominated points.

The solution domains for the sub-problems produced by the conventional branching on the piecewise-linear representation, and the customized branching on the non-dominated points representation are graphically sketched in Figure 4-4. The black lines denote the partitioning enforced by the conventional branching scheme, while the red lines denote the partitioning enforced by the customized branching scheme. The associated arrows indicate the direction of feasibility for the respective sub-problems, and the yellow regions depict the additional non-integral operating points eliminated by the customized branching as compared to the conventional branching.
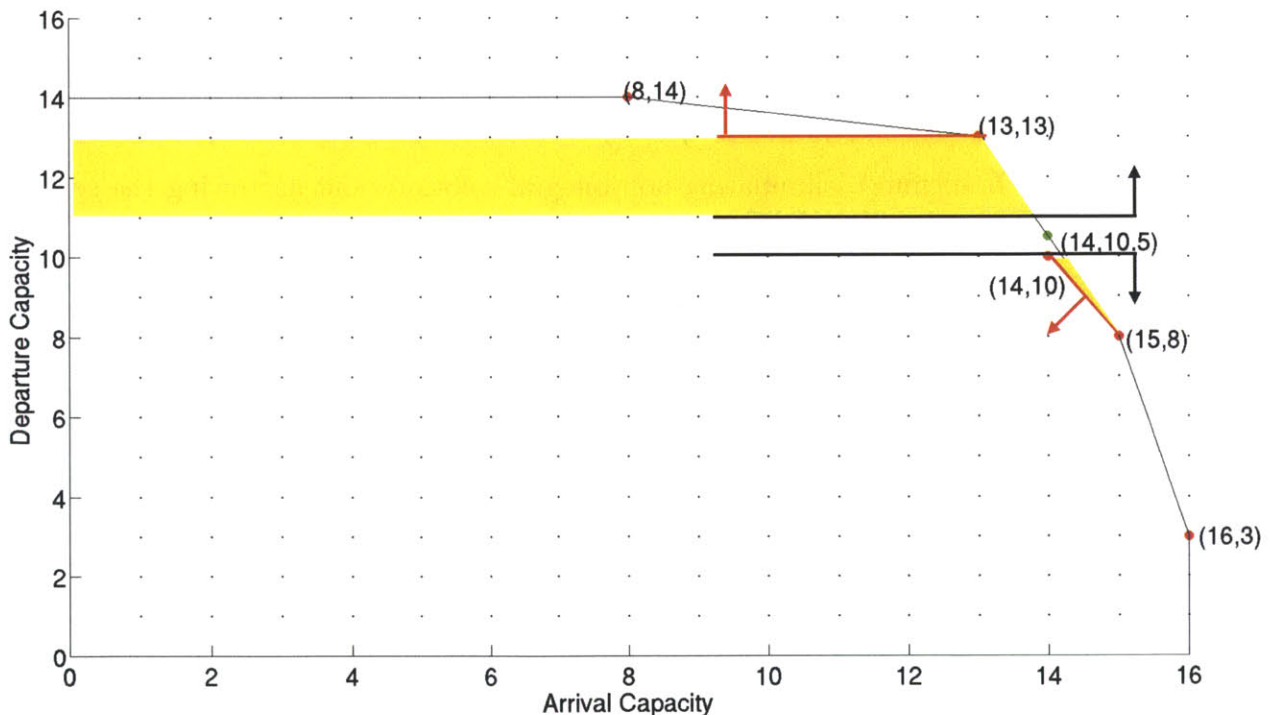
Figure 4-4: Branching schemes on sample capacity envelope.

A general description for the customized branching scheme, based on the proposed ordering of the non-dominated operating points, is also proposed here. Consider a non-integral capacity mix $\phi^i \ \forall i \in \mathbb{ND}_{s,t}$ for some scenario s and time t in the solution for the integrated stochastic formulation (either static or dynamic). Let $L \subset \mathbb{ND}_{s,t}$ be the set of non-dominated points with non-zero values such that $\phi^i > 0 \ \forall i \in L$. Select index $L\_\min = \min_{i \in L} i$, and impose disjunctive branching constraints $\sum_{i \leq L\_min} \phi^i = 1$ and $\sum_{i > L\_min} \phi^i = 1$. The presence of binary variables for the non-dominated points permits further improvization to the branching schemes that are not delved into here. Note that the customized branching schemes are particularly effective when the non-dominated points do not all lie on the boundary of the capacity envelope, such as the point $\{14, 10\}$ in Figure 4-3.

We believe that these non-dominated operating points provide a more natural way of modeling the capacity envelope in integrated stochastic formulations, and in conjunction with the customized branching scheme, are guaranteed to produce some improvement in solution times compared to the piecewise-linear representation of capacity envelope. The magnitude of this run-time improvement is likely to grow with the size of the stochastic ground-holding problem.

Section 4.6 describes experimental case studies that compare ground delay allocations between the integrated stochastic and the arrival-prioritizing frameworks.

4.6   Case Studies

The goal of this section is to quantify and characterize the improvements generated by the integrated stochastic framework for two features of ground-hold allocation: the expected cost and equity. Trends in the magnitude of improvements, in relation to key GDP inputs like arrival/departure demand and capacity forecasts, are assessed for both static and dynamic models.

In the first set of case studies (Section 4.6.1), the GDP demand and capacity inputs are hypothetical, while in Section 4.6.2, demand is based on airline schedules during observed GDPs at LGA, and the capacity envelope measure is as estimated in Chapter 2 using historical ASPM data.

4.6.1 Hypothetical case studies

The experimental design for the hypothetical studies, including the base data and the test parameter that is varied to generate the GDP cases, is presented here. Described next are the chosen performance measures and the corresponding results across all GDP cases for each of the two attributes of ground-holding allocation mentioned above.

4.6.1.1 Baseline data

The key inputs for the stochastic ground-holding models comprise of capacity forecasts, scheduled demand and unit costs for ground and airborne delay. The capacity forecast is represented using a scenario tree. In all the GDP cases considered in this study, the airport capacity envelope is assumed to take one of two forms: a regular form during good weather, and a diminished form during poor weather. The magnitude of capacity decrease during poor weather is quantified using the parameter θ in the following manner:

Given the linear pieces constituting the regular capacity envelope are $\alpha^i A^{\mathrm{cap}} + \beta^i D^{\mathrm{cap}} \leq \gamma^i$, $\forall i \in E$, the corresponding linear pieces for the diminished capacity envelope would be: $\alpha^i A^{\mathrm{cap}} + \beta^i D^{\mathrm{cap}} \leq \theta\gamma^i$, $\forall i \in E$. The parameter θ is termed the capacity decrease ratio. For the hypothetical case studies, the capacity envelope presented in Figure 4-3 was used as the regular capacity envelope, and the capacity decrease ratio was taken to be 0.5. The diminished capacity envelope is shown in Figure 4-5.
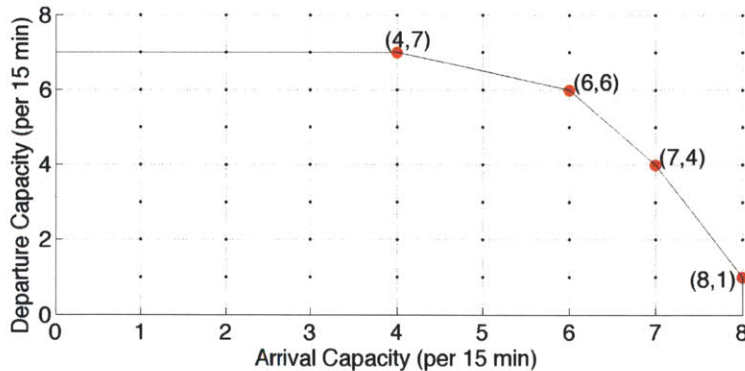
Figure 4-5: Hypothetical diminished capacity envelope ($\theta = 0.5$).

All scenarios begin with capacity at the diminished state, and differ from each other based on the time interval after which capacity improves to the regular state. Accordingly,

116

in the scenario tree, a scenario diverges out from the main branch at the time interval when the capacity improves. The GDP planning horizon is set at 10 time intervals, and the corresponding scenario tree is depicted in Figure 4-6. The diminished capacity states are highlighted in red, and the regular capacity states in green. The scenarios are labelled as $s \in \{1, .., 10\}$, where s represents the duration of diminished capacity.
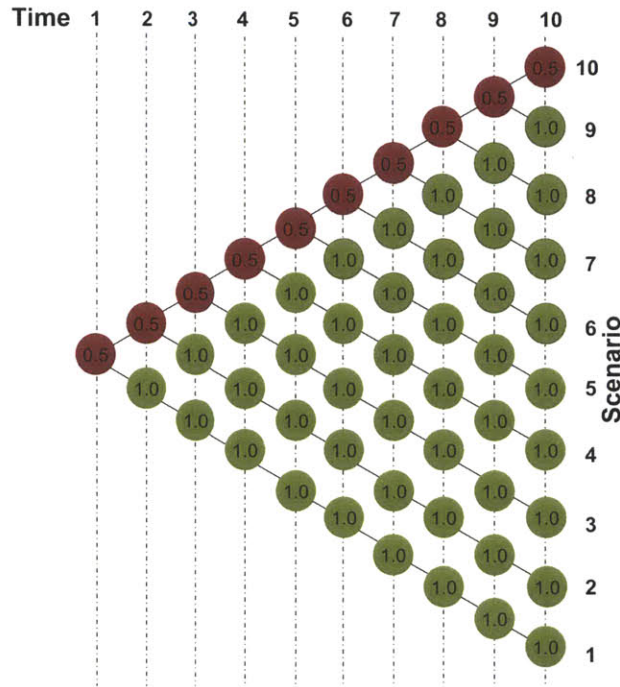
Figure 4-6: Scenario tree format for hypothetical case studies. The red nodes denote the diminished capacity states ($\theta = 0.5$) and the green nodes denote regular capacity ($\theta = 1$).

The following vector of arrival and departure demands was assumed for the 10 time intervals: $(16, 3), (16, 3), (15, 7), (16, 3), (15, 7), (16, 3), (16, 3), (15, 7), (8, 12), (5, 5)$. The flight durations for the arrivals ranged from 2 to 5 time intervals in length. The demand has a high proportion of arrivals, and is feasible for the regular capacity envelope in every time interval. The unit ground and airborne delay costs are set as 0.5 and 2.5 respectively, implying that the ratio of the unit airborne delay cost to unit ground delay cost is greater than the steepest slope of the assumed capacity envelopes. Together with the large fraction of arrivals in the scheduled demand, these GDP inputs are most likely to justify a pre-determined arrival prioritization policy in capacity allocation. If arrival prioritization is suboptimal in this setting, it would offer strong evidence for the benefits of the integrated stochastic framework.

4.6.1.2 Test parameters

The probability distribution across the scenarios shown in Figure 4-6 is varied to generated different GDP cases. We produce a continuum of "expected duration of diminished capacity" values through the following scenario probability distributions, p_s across each of the 10 capacity scenarios:

$$\forall t \in 1, .., 10: \quad p_s^t = \begin{cases} \frac{(1-0.01(10-t))}{t}, & \forall s \in \{1, ..t\} \\ 0.01, & \forall s \in \{t+1, .., 10\} \end{cases} \tag{4.47}$$

$$\forall t \in 1, .., 9: \quad p_s^t = \begin{cases} 0.01, & \forall s \in \{1, ..t\} \\ \frac{(1-0.01t)}{(10-t)}, & \forall s \in \{t+1, .., 10\} \end{cases} \tag{4.48}$$

Note that across these 19 scenario probability distributions, the expected duration of diminished capacity approximately follows the arithmetic sequence $\{1, 1.5, 2.0, ..., 10\}$.

4.6.1.2.1 Delay costs:
This segment details the experimental study of delay costs generated by the integrated stochastic models in comparison to the arrival-prioritizing stochastic models for the above range of GDP cases.

4.6.1.2.2 Performance Measures:
The performance measures chosen for this comparative analysis are the expected system delay cost (C_{sys}), and its individual components: the expected arrival ground (C_{ag}), expected arrival airborne (C_{aa}) and expected departure ground delay costs (C_{dg}). The focus is on the percentage improvements generated in the above delay cost values by the integrated stochastic approach over the arrival-prioritization approach, and the trends exhibited with respect to the expected duration of diminished capacity. In the arrival-prioritization approach, the optimal arrival delay cost $(C_{\text{ag}} + C_{\text{aa}})$ is first obtained from the stochastic model assuming availability of maximum arrival capacity, and the departure ground-holding cost (C_{dg}) is then computed using the residual airport capacity.

118

4.6.1.3 Results:

We first focus on instances of optimal capacity mixes for selected GDP cases to demonstrate the principles of capacity sharing discussed in Section 4.3. Consider two GDP cases with expected duration of diminished capacity of 2 and 9 time-steps, respectively. The first corresponds to a GDP with a high capacity forecast (expected early improvement), while the second corresponds to one with a low capacity forecast (expected late improvement).

Figures 4-7 and 4-8 plot the optimal capacity mix patterns over the entire scenario tree for the integrated static and arrival-prioritizing static models when applied to these two GDP cases. The labels on each node provide the allocated arrival capacity for that scenario and time-step. The green nodes denote regular capacity states, while the red ones correspond to diminished capacity states. Note that for the diminished capacity envelope (Figure 4-5), the non-dominated operating point with maximum arrival capacity is $(8, 1)$, while the operating points with maximum number of operations are $(6, 6)$ and $(7, 5)$. The equivalent points for the regular capacity envelope (Figure 4-3) are $(16, 3)$ and $(13, 13)$.

(a) Arrival-prioritizing static   (b) Integrated static
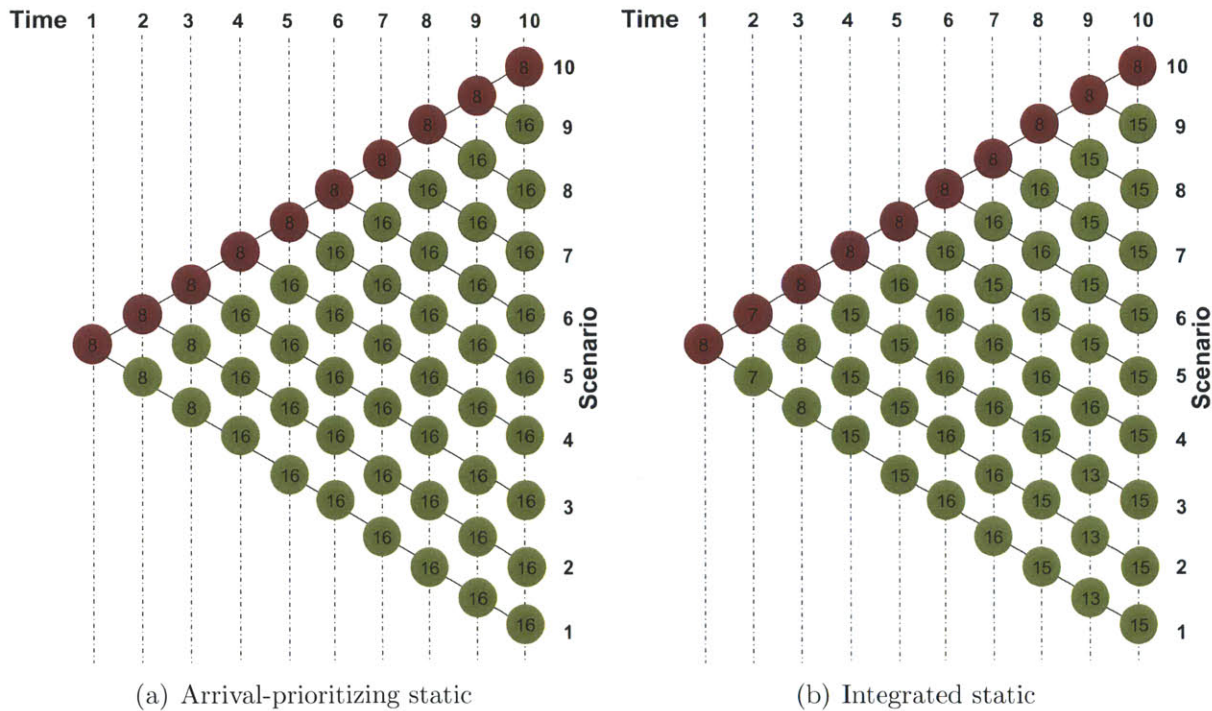
Figure 4-7: Optimal capacity mix for arrival-prioritized and integrated static models with expected duration of diminished capacity of 2 time-steps. The labels on the nodes refer to the allocated arrival capacity of the optimal solution.

(a) Arrival-prioritizing static

(b) Integrated static
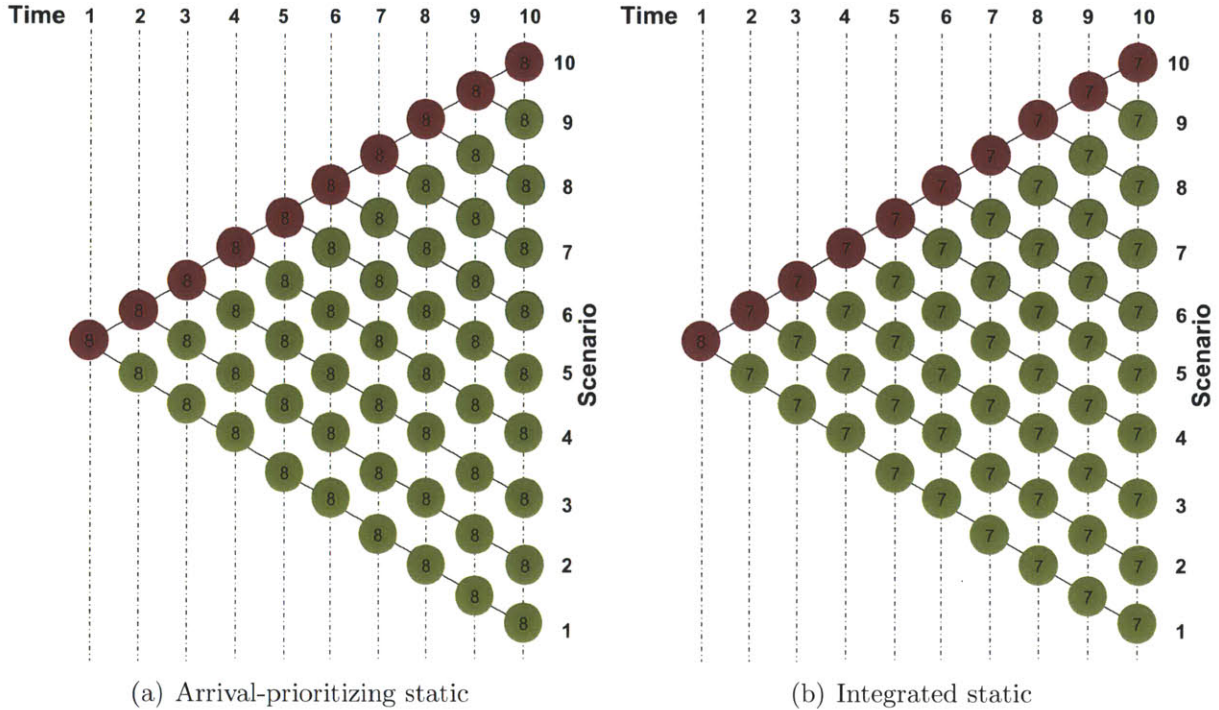
Figure 4-8: Optimal capacity mix for arrival-prioritized and integrated static models, with expected duration of diminished capacity of 9 time-steps. The labels on the nodes refer to the allocated arrival capacity of the optimal solution.

For the GDP case with high capacity forecast (expected duration of diminished capacity $= 2$), the diminished capacity states are typically operated at maximum arrival capacity (i.e., 8) for both the integrated and arrival-prioritized solutions. However, for the normal capacity states, there is a noticeable shift from the maximum arrival capacity (i.e., 16) in the arrival-prioritized solution towards the maximum operational capacity $(13, 13)$ in the integrated solution. Due to the relatively lower magnitude of departure demand, most of the regular capacity states are not required to operate at $(13, 13)$ to serve the departure demand fully.

For the GDP case with low capacity forecast (expected duration of diminished capacity $= 9$), we notice that the shift towards the maximum operational capacity occurs in the diminished capacity states as well (from an allocated arrival capacity of 8 for the arrival-prioritized solution to 7 for the integrated solution). Unlike the GDP case with high capacity forecast, the arrival ground-hold allocation for this GDP case is more conservative in an attempt to prevent airborne queues during the more likely low capacity scenarios. This

120

results in under-utilization of arrival capacity during the regular capacity states.

The results for these two GDP cases indicate that it can be beneficial to incorporate departures into the ground-holding problem, thereby supporting the conceptual arguments presented in Section 4.3.

Figures 4-9 and 4-10 present the percentage changes in delay cost components produced by the integrated approach relative to the arrival-prioritizing approach across the range of generated GDP cases for the static and the dynamic models respectively. The percentage changes in the expected system delay costs (C_{sys}) is plotted in these figures, along with its three components: arrival ground delay (C_{ag}), arrival airborne delay (C_{aa}), and departure ground delay (C_{dg}). As shown in the figures, the integrated versions of both static and dynamic models deliver consistent improvements in system delay costs across the entire range of GDP cases. These improvements come at the expense of arrival ground delay costs, indicating some arrival capacity is transferred to departures in the integrated model solutions.
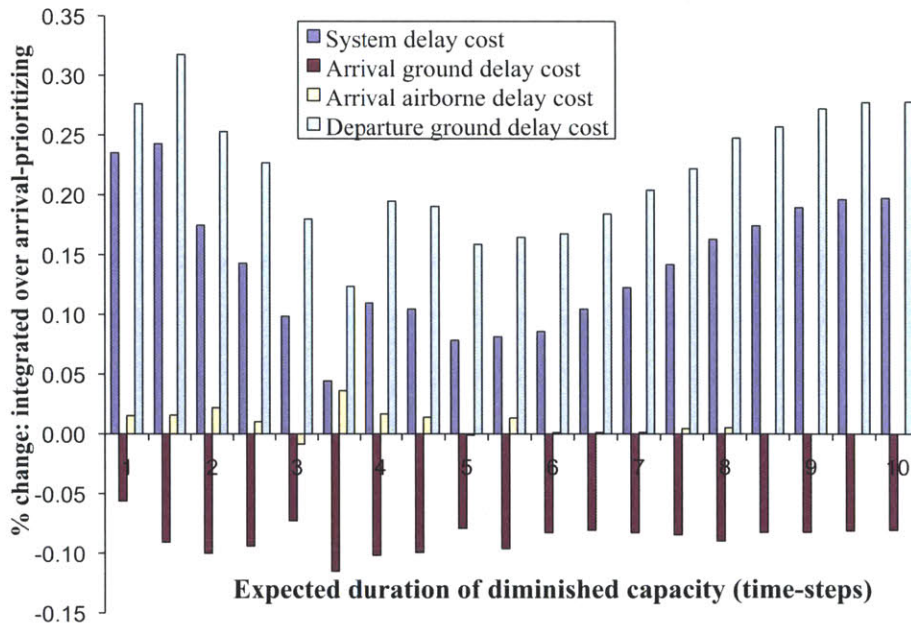
Figure 4-9: Delay cost improvements for integrated framework over arrivals-prioritizing framework (Static stochastic model). Negative values indicate an increase in cost.

An interesting trend observed in the system delay cost improvements for both static and dynamic models is its convexity with respect to the expected duration of diminished

121
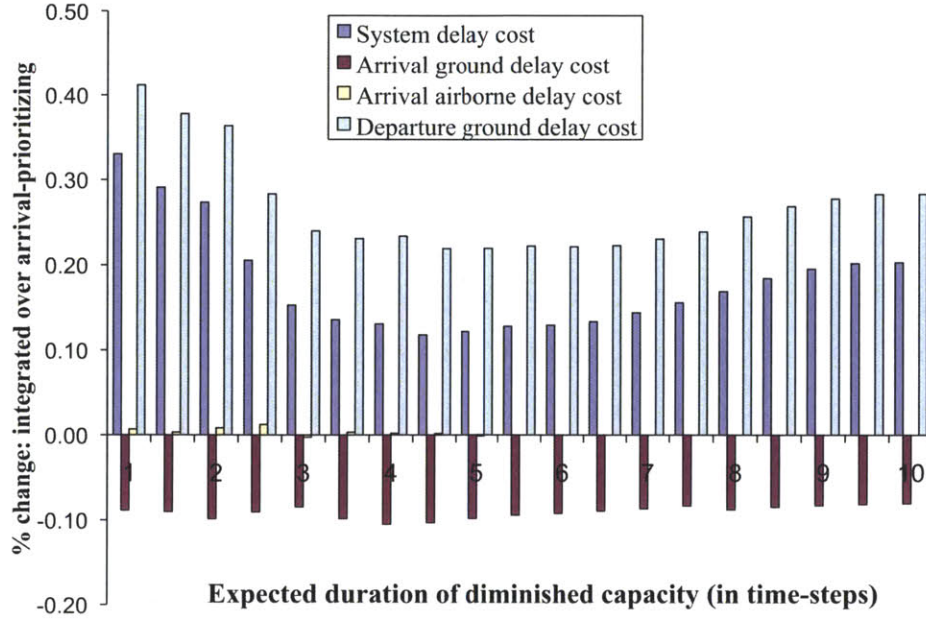
Figure 4-10: Delay cost improvements for integrated framework over arrivals-prioritizing framework (Dynamic stochastic model). Negative values indicate an increase in cost.

capacity. The percentage improvement is high at the extremes, and low for medium values of this parameter, and is consistent with the percentage improvement in departure ground delay costs. This trend can be explained using the results for optimal capacity mix presented in Figures 4-7 and 4-8. This explanation applies to both static and dynamic model results.

For GDP cases with a short expected duration of diminished capacity (Figure 4-7), capacity sharing is limited to regular capacity states in the solution to the integrated approach. There is greater improvement in departure ground delay costs for the scenarios with lower number of diminished capacity states, thereby resulting in higher expected departure cost improvements when such scenarios are more likely. On the other hand, for GDP cases with a long expected duration of diminished capacity (Figure 4-8), capacity sharing in the integrated model solution is observed for diminished capacity states as well. Due to conservative nature of arrival delay allocation for such GDP cases, high departure capacity is available during the regular capacity states even in the arrival-prioritizing solution. The departure delay cost improvements produced by the integrated solution are concentrated in scenarios with prolonged diminished capacity, thereby resulting in higher expected departure cost improvements when such scenarios are more likely. The above-described convexity in delay

cost improvements is specific to the idealized format of the GDP inputs, particularly the predominance of arrival demand.

Lastly, the percentage improvements in system delay costs are typically higher for the dynamic model (Figure 4-10) than for the static model (Figure 4-9). This reflects the ability of the dynamic model to better utilize available airport capacity through dynamic revisions of ground delay allocation for both arrivals and departures, which results in a lower increase in arrival delay costs for the integrated dynamic model compared to the integrated static model.

4.6.1.4 Equity

In this section, the equity of the ground-holding allocation generated by the integrated stochastic models is examined in comparison to the arrival-prioritizing stochastic models for the above range of GDP cases.

4.6.1.4.1 Performance Measures: The concept of equity of delay allocation in air traffic management has grown in significance following the advent of Collaborative Decision-Making (CDM). Standard interpretation of equity in literature relates to the preservation of original order of flight schedule following delay allocation [18, 75, 5]. This principle is explicitly enforced in the Ration-by-schedule allocation (RBS) scheme adopted for arrival slot allocation in modern GDPs, which ensures the original arrival schedule is preserved following delay allocation. Metrics proposed in recent work to quantify the inequity of delay allocation schemes typically represent a measure of the deviation from original schedule [12, 8]. In case of the stochastic ground-holding models, the original departure schedule is always preserved in the final solution. Hence, we focus on inequity in arrival delay allocation.

Bertsimas and Gupta (2009) [12] suggest measuring the number of pair-wise reversals in the delayed arrival schedule relative to the original schedule. Barnhart et. al. [8] proposes measuring the incremental delay experienced by each arrival in the delayed schedule, in excess of that experienced by the flight landing in its original position in the scheduled arrival order, and aggregating this measure across all arrivals. Let us denote these metrics as E_{Gupta} and E_{Barnhart}. While both metrics are defined for deterministic capacity settings,

they can be easily adapted to a stochastic setting by computing their expected measure across all scenarios: $E_{\text{Gupta}} = \sum_{s \in S} p_s E_{\text{Gupta}}^s$, $E_{\text{Barnhart}} = \sum_{s \in S} p_s E_{\text{Barnhart}}^s$.

As described in Section 4.4, the dynamic stochastic model inherently gives rise to inequity in the arrival ground-hold allocation by favoring long-haul over short-haul flights. The above metrics are used to evaluate the differences in arrival delay inequity between the integrated dynamic solution and the arrival-prioritizing dynamic solution.

4.6.1.5 Results

Figures 4-11 and 4-12 respectively plot the computed values for the metrics proposed by Bertsimas and Gupta (2009) [12] and Barnhart et. al. (2009) [8] across the spectrum of GDP cases for the integrated and arrival-prioritizing models.
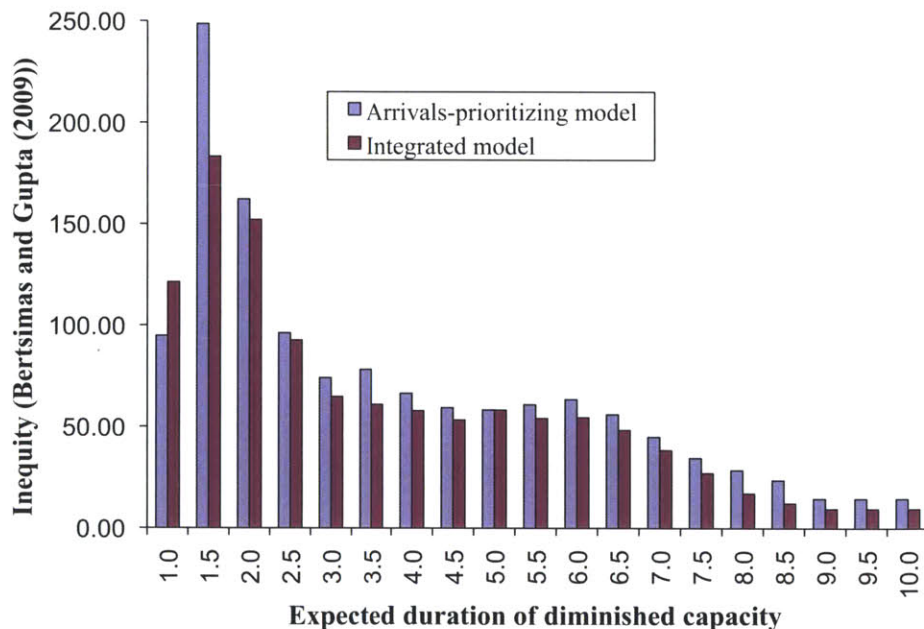
Figure 4-11: Inequity in arrival delay allocation for integrated and arrival-prioritizing dynamic models, measured using the Bertsimas and Gupta (2009) metric [12].

Given that a lower value for each metric implies an improvement in allocation equity, both these figures demonstrate the more equitable behavior of the integrated dynamic model for most GDP cases. This observation can be explained by the basic principles driving the inequity of the dynamic model. Since the dynamic model revises ground-hold decisions at every time-step, the decision for a short-haul arrival can take place at a later time, when com-

Figure 4-12: Inequity in arrival delay allocation for integrated and arrival-prioritizing dynamic models, measured using the Barnhart et al. (2009) metric [8].

pared to a long-haul arrival scheduled to land at the same time. Therefore, there is greater amount of information available on capacity scenario materialization when rescheduling a short-haul flight. A short-haul flight is also more responsive to evolving capacity information, as it can make quicker use of additional capacity whenever it materializes. These factors are responsible for short-haul arrivals absorbing most of the capacity uncertainties in the dynamic model allocation, be it larger delays for scenarios with prolonged low capacity or quick departures to utilize a sudden capacity increase. This feature results in the disparity between short-haul and long-haul arrival allocations. In an integrated framework, departures are potentially best suited to absorb the capacity uncertainties, thereby mitigating some of the variability in the short-haul arrivals' allocations and the resulting disparity with the long-haul arrivals. In this manner, an integrated framework for dynamic delay allocation can potentially help reduce the inequity between the long-haul and short-haul arrivals, as evidenced in Figures 4-11 and 4-12 for the hypothetical case studies.

125

4.6.2 Real-world case studies

In this section, the findings from the hypothetical case studies are validated using real-world data from recorded GDPs at LGA airport.

4.6.2.1 Input data

Details on timing and operational throughputs of every initiated GDP at LGA are available in GDP summary files. By matching these times with those in the ASPM database, information on scheduled arrivals and departures during recorded GDPs can be extracted. The GDP recorded on Feb 17, 2006 from 7 am to midnight was selected for this study, and concurrent arrival and departure schedules from ASPM were extracted. Given below are the hourly aggregate arrival and departure demand for the first 10 hours of this GDP. The aggregate arrival schedule matrix (Table 4.2) breaks down the hourly aggregate into different flight duration categories. Flight durations were rounded up to the nearest integer hour.

Hour	1	2	3	4	5	6	7	8	9	10
Demand	13	39	48	39	29	30	35	33	38	37

Table 4.1: Hourly aggregate scheduled departures (0700-1700 on Feb 17, 2006).

Hour	1-hr	2-hr	3-hr	4-hr	5-hr	Total
1	3	1	0	0	0	4
2	5	30	0	0	0	35
3	7	24	5	0	0	36
4	6	15	7	1	0	29
5	5	15	6	3	2	31
6	7	21	8	3	0	39
7	7	17	4	3	2	33
8	6	16	10	2	2	36
9	6	23	10	1	0	40
10	5	13	7	3	4	32

Table 4.2: Hourly aggregate scheduled arrivals (0700-1700 hours on Feb 17, 2006).

Unlike the demand inputs to the hypothetical case studies, there is no significant disparity between arrival and departure demands over the 10-hr time horizon. Due to the lack of reliable data on actual capacity forecasts used during GDPs, we use idealized scenario trees

126

characterized by shrinkage ratio θ as designed for the hypothetical case studies (Figure 4-6). However, in place of the idealized capacity envelope sketched in Figure 4-4, the good-weather (VFR) capacity envelope for LGA as estimated in Chapter 2 using historical ASPM data on airport throughputs was used. As was seen in the hypothetical case studies, a set of scenario probability distributions spanning a range of values for the expected duration of diminished capacity was generated, with the capacity decrease ratio θ set to 0.5 again. The unit costs for ground and airborne delays are also assumed to be 0.5 and 2.5. Figure 4-13 depicts the good-weather and diminished capacity envelopes, along with the operational demand points over the 10-hr time horizon. Note that the envelopes are scaled up to 1-hour counts for the purpose of this case study.
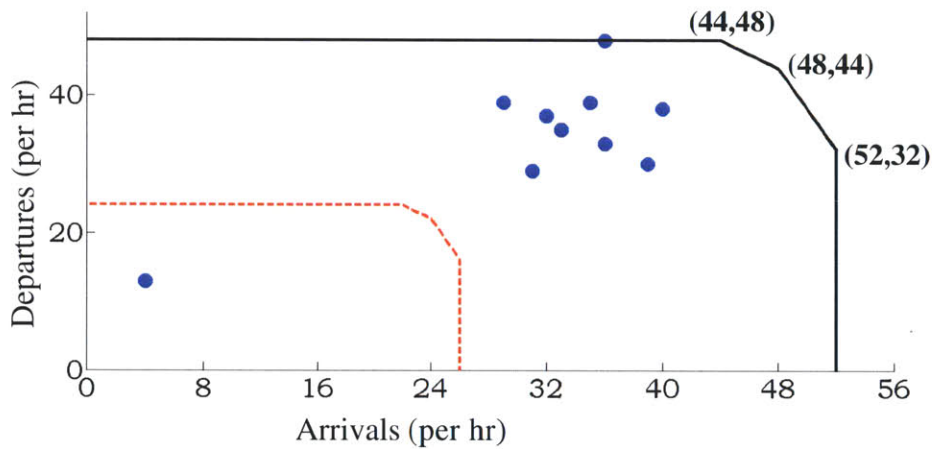
Figure 4-13: Capacity envelopes and demand points for LGA case study.

The performance measures on system delay costs and equity were computed for the GDP cases with the real-world inputs of scheduled demand, and the regular capacity envelope at LGA.

4.6.2.2   Results

Figures 4-14 and 4-15 present the percentage changes in delay cost components produced by the integrated approach relative to the arrival-prioritizing approach across the range of GDP cases involving real-world inputs, for the static and dynamic models respectively.

127
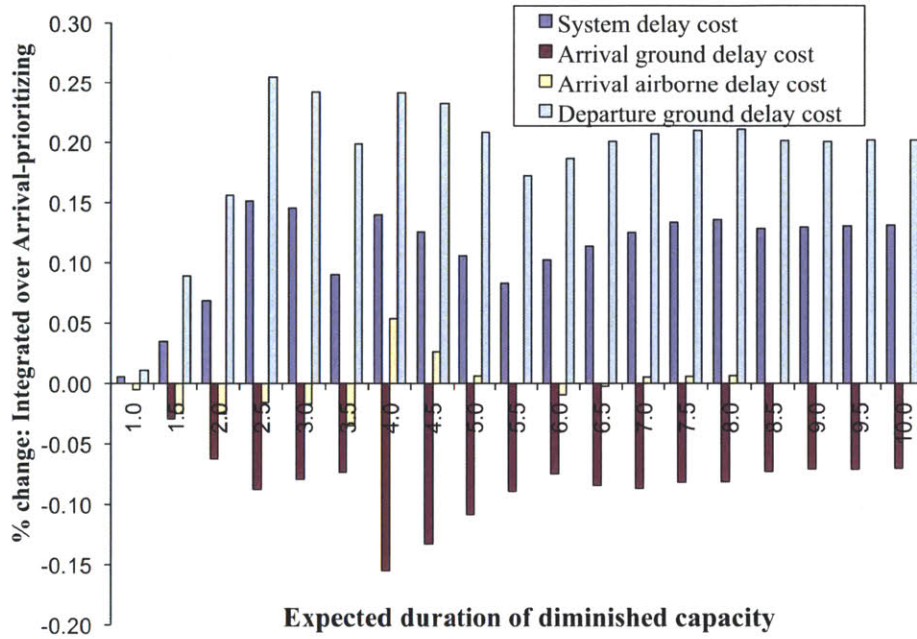
Figure 4-14: Delay cost improvements for integrated framework over arrivals-prioritizing framework on real-world case studies (Static stochastic model). Negative values indicate an increase in cost.
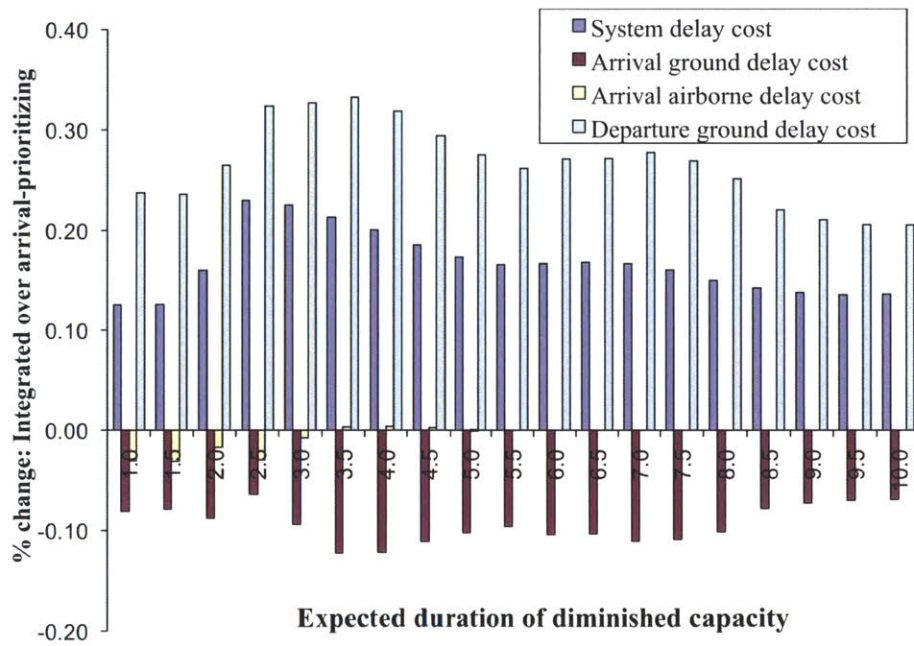
Figure 4-15: Delay cost improvements for integrated framework over arrivals-prioritizing framework on real-world case studies (Dynamic stochastic model). Negative values indicate an increase in cost.

The trends observed in these results are consistent with those observed for the hypothetical case studies in Figures 4-9 and 4-10. The improvements in system delay costs come at the expense of an increase in arrival ground delay costs, and are reasonable in magnitude

128

for every GDP case. Also, the integrated dynamic model typically generates greater system delay improvements than its static counterpart. A key contrast to the hypothetical case study is the absence of a convex relationship between the system delay cost improvements and the expected duration of diminished capacity. This observation can be attributed to the lack of arrival predominance in the scheduled demand in the real-world case studies.

Figures 4-16 and 4-17 depict the computed values of the Bertsimas and Gupta (2009), and Barnhart et al. (2009) equity metrics for the optimal arrival ground delay allocations, from the integrated and arrival-prioritizing dynamic stochastic models. The integrated dynamic model produces an improvement in both equity metrics through almost the entire spectrum of real-world GDP cases. These results further support the hypothesis that an integrated framework for stochastic ground-holding typically facilitates a more equitable allocation of delays across scheduled arrivals, by virtue of balanced capacity sharing between arrivals and departures.
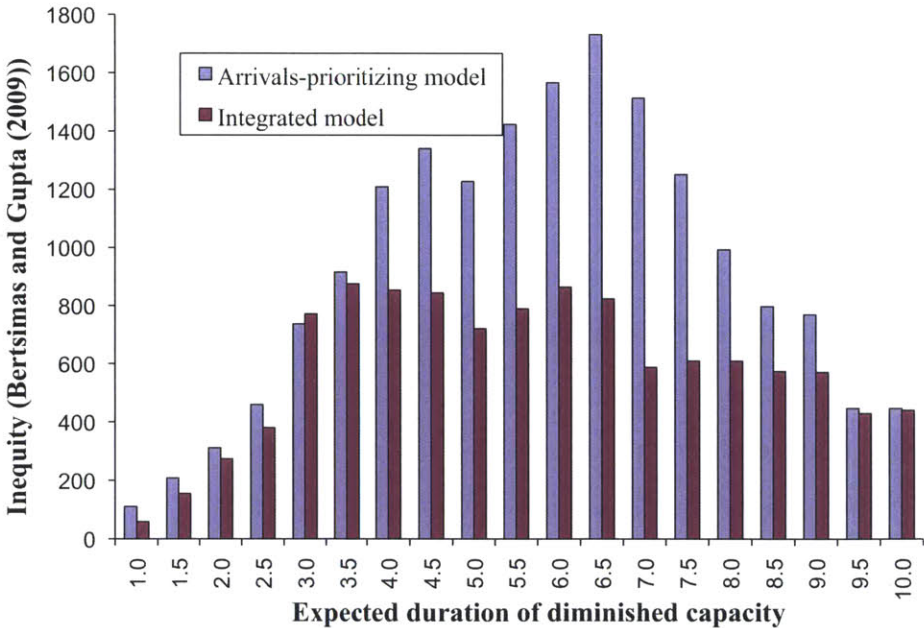
Figure 4-16: Inequity in arrival delay allocation for integrated and arrival-prioritizing dynamic models, measured using the Bertsimas and Gupta (2009) metric [12] for real-world GDP cases.

A key objective of this thesis is to study the implementation of CDM mechanisms in conjunction with the stochastic ground-holding models discussed in this chapter. This goal requires the development of stochastic equivalents of the incumbent CDM mechanisms that
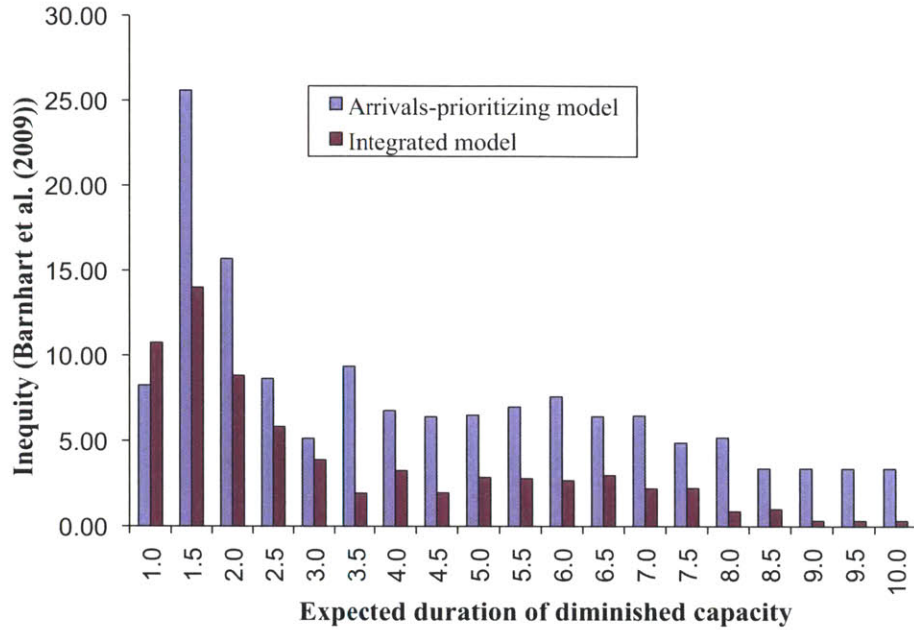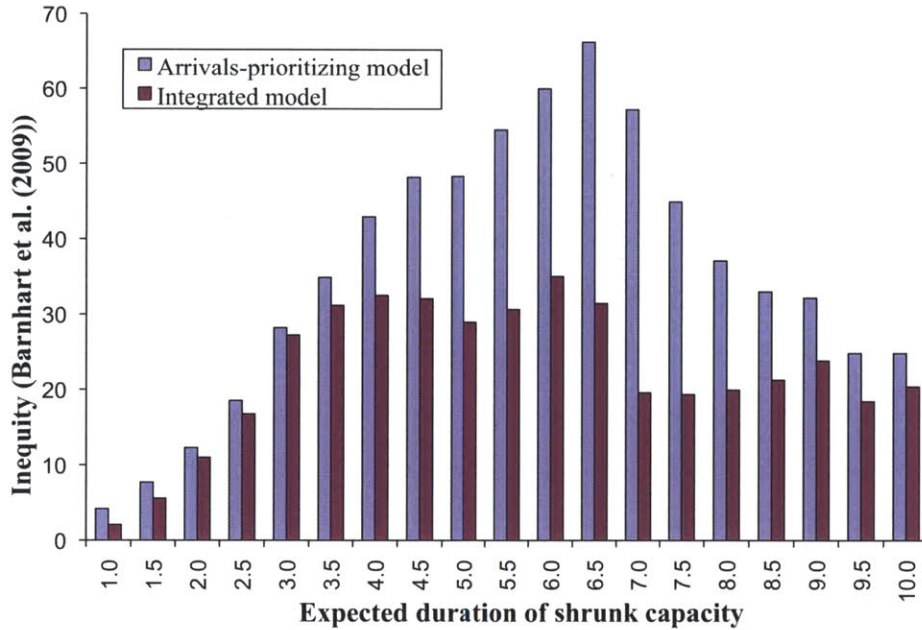
Figure 4-17: Inequity in arrival delay allocation for integrated and arrival-prioritizing dynamic models, measured using the Barnhart et al. (2009) metric [8] for real-world GDP cases.

can be applied to the slot allocation solutions obtained from stochastic ground-holding models. Before delving into such an investigation, the issue of compatibility of the solutions of various stochastic ground-holding models with CDM mechanisms needs to be addressed. The amount of flexibility offered to the airlines for substituting and canceling slots depends on whether a static or a dynamic stochastic model is used to determine the slot allocation. This factor is a driver of the tradeoffs between static and dynamic stochastic models, as discussed in the following chapter. These tradeoffs motivate the design of a hybridized stochastic model that combines the favorable features of static and dynamic models. The development of such a hybrid stochastic formulation is the focus of the next chapter.

Chapter 5

Hybrid Stochastic Ground-Holding Model

5.1 Introduction

This chapter focuses on the relative advantages and disadvantages of static and dynamic stochastic ground-holding models within the context of Collaborative Decision-Making (CDM). Both approaches are single-airport models that allocate ground delays to scheduled arrivals when future capacity is uncertain. They are formulated as integer stochastic programs that minimize the expected sum of ground and airborne delay costs across a finite set of possible capacity scenarios with specified probabilities. As seen in Chapter 4, the static model of Richetta and Odoni (1993) [67] is a single-stage integer stochastic program, while the dynamic model of Mukherjee and Hansen (2007) [55] is a multi-stage integer stochastic program.

During present-day GDPs, arrival slots allocated by the ground-holding model are adjusted further with direct participation from the airlines. The two CDM mechanisms that enable these adjustments are designed to enhance system efficiency (Chang et. al., 2001) [18]. The first one, *intra-airline slot substitution*, allows airlines to swap slots allotted to them in order to to satisfy flight-specific delay preferences. Through the second mechanism, *compression*, airlines can exchange an assigned earlier slot in the event of mechanical delays or a cancellation, for a later slot.

131

Deterministic and stochastic ground-holding models both assume homogeneous unit costs for ground and airborne delay across all flights. It is this assumption that motivates the design of intra-airline slot substitution mechanism as means of improving system efficiency. Since flight-specific delay costs are usually private to the airlines, the slot substitution mechanism allows airlines to improve their internal delay costs by reassigning their flights among their own slots. Airlines are not required to reveal their delay costs, thereby alleviating concerns of incentive-compatibility (the guarantee of truthful revelation), as well as real-time tractability.

By nature of their respective formulations, the static stochastic model determines ground-holding for all flights in the GDP at the outset, while the dynamic stochastic model permits revision of ground-holding decisions based on the latest information on scenario materialization. The dynamic model will therefore always achieve lower optimal system delay costs, assuming homogeneous unit costs. However, dynamic stochastic ground-holding formulation relies upon the travel durations of individual flights. Section 5.2 illustrates how this property of the dynamic solution limits slot substitution, as slots assigned to a short-haul flight may be incompatible for longer flights over the set of possible capacity scenarios. This feature produces a trade-off between potentially greater pre-CDM delay cost reduction and potentially lower CDM-induced reduction in airline delay costs for the dynamic model, when compared to the static model.

This chapter integrates the favorable features of the static and dynamic models into a *hybrid stochastic ground-holding model*. Similar to the dynamic model, the hybrid model uses the latest information on capacity scenario materialization, and yet eliminates the dependence of its ground-holding solution on flight duration. This chapter also establishes two useful results on the tractability of the integer stochastic formulation of the hybrid model, when reasonably realistic conditions are satisfied by its input parameters. The first result is applicable for marginally non-decreasing ground-hold cost functions, under which condition the linear relaxation of static model formulation was recently shown to yield integral solutions by Kotnyek and Richetta (2006) [49]. Given the same condition, we prove that integrality of the hybrid formulation can be guaranteed by imposing integrality requirements on a subset of the decision variables. We also prove that an added condition on the structure

of the capacity scenario tree establishes a stronger version of the first result.

This chapter uses the original arrivals-based formulations of stochastic ground-holding models, as opposed to the integrated versions presented in the Chapter 4. This restriction is for clarity, and the central motivation behind the development of the hybrid stochastic model will be retained in an integrated framework as well.

5.2 Comparison of Static and Dynamic Stochastic Models

The static model ((4.18)-(4.21)) is formulated as a single-stage stochastic integer program, and the first-stage decision variables are the ground-holding values for flights scheduled to arrive within the GDP time window $X_{t,t+n}$. By design, these ground-holds are the same for every capacity scenario $s \in S$ that can potentially materialize.

The dynamic model ((4.31)-(4.35)), on the other hand, is a multi-stage stochastic integer program with disaggregated flight-specific decision variables $X_{f,t}^s$. This variable represents the ground-holding decision for flight f at a stage jointly defined by time-step t and capacity scenario s, and can be subjected to dynamic recourse while satisfying the non-anticipativity constraints (4.34).

As seen in Chapter 4, a multi-stage dynamic stochastic model makes better use of evolving information on capacity scenario materialization, as compared a single-stage static model. It would therefore yield a lower optimal system delay cost for the original (pre-CDM) stochastic ground-holding allocation. The above statement is only valid under the assumption of homogeneous delay costs across flights. In actual GDP settings, delay cost functions may vary among flights for operating airlines, but we expect that the dynamic model will still have lower pre-CDM delay costs.

While it has lower pre-CDM delay costs, the dynamic model differentiates between flights based on their durations, thereby limiting slot substitution options under CDM. This mechanism, termed intra-airline slot substitution, allows an airline to redistribute arrival times amongst its flights to prioritize critical flights. The slot assigned to a flight by the dynamic

model could feature different arrival times under different scenarios depending on its duration. As a consequence, slots allotted to flights of different durations cannot always be feasible swapped across all scenarios, as demonstrated in the following example.

5.2.1 Static and dynamic stochastic ground-holding problem examples

Consider a GDP with a 6 time-step capacity scenario tree as given in Figure 5-1, and arrival demand schedule featuring two flights A and B, as given in Table 5.1.
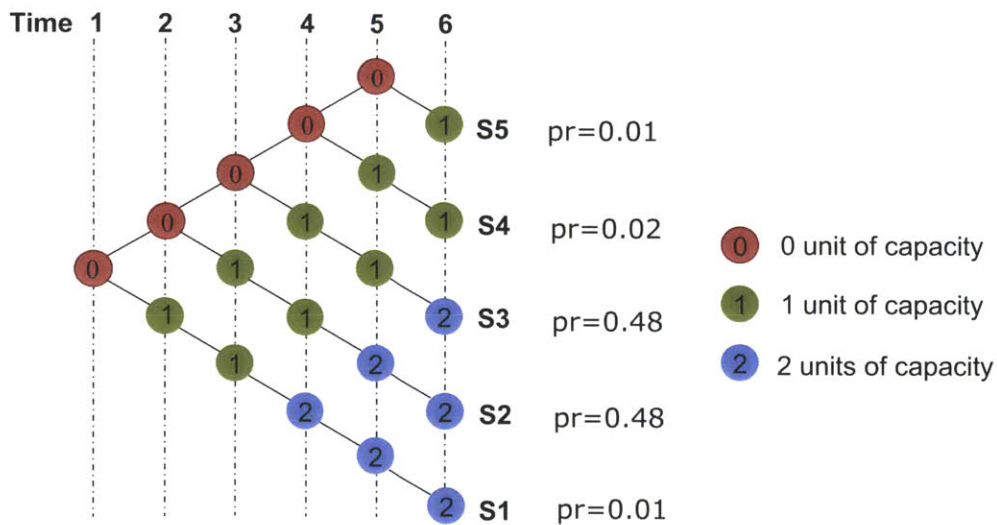
Figure 5-1: Capacity scenario tree for illustrative example

Flight ID	Departure time	Arrival time	Flight duration
A	1	3	2
B	3	4	1

Table 5.1: Arrival demand schedule for a hypothetical example.

Assuming the homogeneous unit ground and airborne delay costs for both flights to be 0.5 and 2.5 respectively, the ground-holding solutions to the static and dynamic models applied upon this example are provided in Tables 5.2 and 5.3 respectively. As expected, the delay cost amount for the dynamic solution is lower, illustrating the dynamic model's pre-CDM advantage.

134

Flight ID		S1	S2	S3	S4	S5
A	Ground-hold	1	1	1	1	1
	Departure time	2	2	2	2	2
	Arrival time	4	4	4	4	4
B	Ground-hold	1	1	1	1	1
	Departure time	4	4	4	4	4
	Arrival time	5	5	5	5	5

Table 5.2: Solution to static stochastic ground-holding problem. Optimal (pre-CDM) delay cost = 1.2.

Flight ID		S1	S2	S3	S4	S5
A	Ground-hold	1	1	1	1	1
	Departure time	2	2	2	2	2
	Arrival time	4	4	4	4	4
B	Ground-hold	0	1	1	2	3
	Departure time	3	4	4	5	6
	Arrival time	4	5	5	6	7

Table 5.3: Solution to dynamic stochastic ground-holding problem. Optimal (pre-CDM) delay cost = 1.065.

If the same airline operates flights A and B, it is possible that the airline would like to swap the scenario-based arrival slots allotted to these two flights. Such a situation might arise if the airline considers flight B to be more delay-sensitive than flight A, perhaps owing to aircraft or passenger connectivity.

Flight ID		S1	S2	S3	S4	S5
A	Departure time	3	3	3	3	3
	Arrival time	5	5	5	5	5
B	Departure time	3	3	3	3	3
	Arrival time	4	4	4	4	4

Table 5.4: Solution to static stochastic ground-holding, post-slot substitution.

Flight ID		S1	S2	S3	S4	S5
A	Departure time	2	3	3	4	5
	Arrival time	4	5	5	6	7
B	Departure time	3	3	3	3	3
	Arrival time	4	4	4	4	4

Table 5.5: Solution to dynamic stochastic ground-holding, post-slot substitution.

We observe, using Tables 5.4 and 5.5, that the desired slot substitution is feasible for the static model, but not so for the dynamic model. In Table 5.5, flight A cannot depart on the 3rd hour under scenario S3 and on the 4th hour under scenario S4, since scenarios S3 and S4 are not distinguishable at start of the 3rd hour. This incompatibility arises because A, a 2-hr long flight, cannot feasibly occupy a scenario-based slot originally allotted to B, a 1-hr long flight.

This example supports our earlier hypothesis that the dynamic model might limit the feasibility of slot substitution among flights of different durations. Consequently, the incremental gains in delay costs for airlines through CDM mechanisms is likely to be lower when using the dynamic model. Like the pre-CDM superiority of the dynamic model, the conceptual CDM-related deficiency of the dynamic model solution need not always translate to lower incremental gains compared to the static model (e.g. if airline operates flights of identical duration), but is expected to do so for typical conditions.

Table 5.6 summarizes the above-described trade-offs between the static and the dynamic model with respect to pre-CDM and CDM allocations.

	Static	Dynamic
Pre-CDM optimal system delay cost	Higher	Lower
CDM-induced delay cost decrease	Larger	Smaller

Table 5.6: Typical tradeoffs between Static and Dynamic Stochastic models for pre-CDM and CDM performances.

5.3 Hybrid Stochastic Ground-Holding Model

Section 5.2 described the relative advantages of the static and dynamic stochastic models with respect to pre-CDM and CDM allocations under typical conditions. In particular, the multi-stage stochastic formulation of the dynamic model results in better pre-CDM allocations, while the flight duration-independent allocations of the static model lead to more slot substitution options. This section develops a new model, called the *hybrid stochastic model* that combines these two favorable features. The formulation of the hybrid stochastic model is presented below.

136

$$\text{Minimize} \quad \sum_{s \in S} p_s \left(\sum_{n=0}^{K} C_{g,n} \left(\sum_{t=1}^{T-n} X_{t,t+n}^s \right) + C_a \sum_{t=1}^{T} A_{s,t}^{\text{aq}} \right) \tag{5.1}$$

subject to:

$$\sum_{j=t}^{t+K} X_{t,j}^s = A_t^d, \ \forall t \in \{1,..,T\}, s \in S \tag{5.2}$$

$$A_{s,t}^{\text{aq}} \geq \sum_{j=t-K}^{t} X_{j,t}^s + A_{s,t-1}^{\text{aq}} - A_{s,t}^{\text{cap}}, \ \forall t \in \{1,..,T\}, s \in S \tag{5.3}$$

$$X_{t,j}^{s^1} = X_{t,j}^{s^2}, \ \forall s_1, s_2 \in G(t - \text{max\_dur}) \tag{5.4}$$

$$X_{t,j}^s \in Z^+, \ \forall t, j \in \{1,..,T\}, s \in S$$

Extended notation (in addition to that of the static and dynamic stochastic formulations)

Input

max_dur : Maximum flight duration among all flights considered for rescheduling under GDP

Decision Variables

$X_{t,t+n}^s$: Number of arrivals with original arrival time t rescheduled to arrival time $t + n$ for scenario s

Objective function

(5.1) : Expected sum of arrival ground delay cost and airborne delay cost

Constraints

(5.2) : Arrival demand balance

(5.3) : Arrival queue balance

(5.4) : Non-anticipativity on arrival ground-holding

The hybrid model is a multi-stage stochastic mixed-integer program like the dynamic model, and permits scenario-specific revision of the ground-holding solution, ensuring a lower optimal pre-CDM delay cost compared to static model (assuming homogeneous costs). At any time-step t under capacity scenario s, while the dynamic model assigns ground-holding

for flights held on ground at that time-step using the variable $X_{f,t}^s$, the hybrid model assigns ground-holding for flights originally scheduled to land at time-step $t + \text{max\_dur}$ using the variable $X_{t,t+n}^s$, $\forall n \in 0, .., K$. This feature ensures that slots assigned to two flights of different durations under the hybrid solution can be feasibly exchanged, as demonstrated in the following example.

5.3.1 Hybrid stochastic ground-holding problem example

Table 5.8 presents the ground-holding solution of the hybrid stochastic model for the example in Section 5.2.1. Note that the homogeneously computed delay cost amount for the hybrid solution is lower than that of the static solution, but higher than that of the dynamic solution. Table 5.9 shows the result of swapping the slots assigned to flights A and B under the hybrid solution. This is a feasible slot substitution: Flight A can depart on the 2nd hour for scenario S1 and on the 3rd hour for other scenarios, since S1 would be distinguishable from the other scenarios by the 2nd hour according to the scenario tree in Figure 5-1.

In light of these properties of the hybrid stochastic model, Table 5.10 summarizes the tradeoffs between the three stochastic models, for typical cost structures.

		S1	S2	S3	S4	S5
	Ground-hold	1	1	1	1	1
A	Departure time	2	2	2	2	2
	Arrival time	4	4	4	4	4
	Ground-hold	0	1	1	1	1
B	Departure time	3	4	4	4	4
	Arrival time	4	5	5	5	5

Table 5.8: Solution to hybrid stochastic ground-holding problem. Optimal (pre-CDM) delay cost = 1.195.

		S1	S2	S3	S4	S5
	Departure time	2	3	3	3	3
A	Arrival time	4	5	5	5	5
	Departure time	3	3	3	3	3
B	Arrival time	4	4	4	4	4

Table 5.9: Solution to hybrid stochastic ground-holding problem, after the slots for flight A and flight B are swapped.

	Static	Hybrid	Dynamic
Pre-CDM optimal system delay cost	Highest	Moderate	Lowest
CDM-induced delay cost decrease	Largest	Moderate	Smallest

Table 5.10: Typical tradeoffs between Static, Hybrid and Dynamic Stochastic models for pre-CDM and CDM performances.

5.3.2 Equity of hybrid stochastic model

Like the dynamic stochastic model, the hybrid stochastic formulation may rearrange the original arrival schedule for some capacity scenarios. As noted in Mukherjee and Hansen (2007) [55], this rearrangement is an undesirable source of inequity in ground-hold allocation. However, the key difference between the dynamic and hybrid solution is that the latter is not biased based on flight duration. The dynamic model can potentially delay short-haul arrivals under low capacity scenarios, as these arrivals are more responsive to evolving capacity information. By contrast, the hybrid model would simply favor arrivals scheduled for later time-steps in the GDP, since they could potentially be advanced in the event of an early increase in airport capacity. Since the length of a GDP is known just prior to its initiation, the nature of inequity imposed by the hybrid solution is unsystematic in comparison to the dynamic model, and therefore preferable.

5.4 Properties of Hybrid Stochastic Formulation

This section develops two key results concerning the tractability of the hybrid stochastic model formulation under a fairly general set of conditions. The proofs assume integer demands ($A_t^d \in \mathbb{Z}^+$, $\forall t \in \{1, .., T\}$) and capacities ($A_{s,t}^{\text{cap}} \in \mathbb{Z}^+$, $\forall s \in S$, $\forall t \in \{1, .., T\}$). The proofs rely upon *perturbation analysis* as a device to establish properties of the optimal solutions. The brief description of perturbation analysis is as follows: We consider a possibly nonconforming optimal solution (i.e., an optimal solution that does not maintain the flight ordering the original schedule), and perturb it by an infinitesimal amount in the direction of a conforming solution, while ensuring no increase to the objective function value or violation of constraints. We thereby construct a feasible, conforming optimal solution that can be obtained by employing a continuous sequence of optimality-preserving perturbations to the

non-conforming solution.

5.4.1 Case 1: Integer queue lengths

Given *marginally non-decreasing ground-holding cost coefficients* $(C_{g,n+1} - C_{g,n} \geq C_{g,n} - C_{g,n-1} \; \forall n)$, the hybrid stochastic formulation is guaranteed to have an integral optimum solution if the queue length variables $(A^{\mathrm{aq}}_{s,t}, \; \forall s \in S, t \in \{1,..,T\})$ are constrained to have integer values.

5.4.1.1 Proof

Assume an optimal solution X (with optimal cost Z) for the hybrid formulation satisfying the specified conditions on $C_{g,n}$, and with $A^{\mathrm{aq}}_{s,t} \in \mathbb{Z}^+ \; \forall s \in S, t \in \{1,..,T\}$, but with fractional solution values $X^s_{a,b}$ for some $s \in \{1,..,S\}; a,b \in \{1,..,T\}$. We now describe an algorithm that converts this solution into a fully integral solution, through a sequence of perturbations that do not increase the optimal cost value.

5.4.1.1.1 Algorithm

1. Amongst the fractional values, let $i = \min\limits_{a: X^s_{a,b} \notin \mathbb{Z}^+ \; \forall s \in S} a$, and $j = \min\limits_{b: X^s_{i,b} \notin \mathbb{Z}^+ \; \forall s \in \{1,..,S\}} b$. By manner of selection of indices i and j, there does not exist $p < i$ such that $X^s_{p,j} \notin \mathbb{Z}^+$ for any $s \in S$.

2. Since $\sum\limits_{b=i}^{i+K} X^s_{i,b} = A^d_i$ and $A^d_i \in \mathbb{Z}^+$, we know that if $X^s_{i,j} \notin \mathbb{Z}^+$ and i is selected as described in Step 1, there exists $b > j$ such that $X^s_{i,b} \notin \mathbb{Z}^+$. The above is true because the reallocations of arrival demand originally scheduled at time i need to add up to an integer value. Let $k = \min\limits_{X^s_{i,b} \notin \mathbb{Z}^+ \mathrm{and} X^s_{i,j} \notin \mathbb{Z}^+} b$.

3. We know, by the non-anticipativity constraints (5.4), that $X^s_{i,b} \; \forall b \in i,..,i+K$ are the same for all $s \in G(i - \mathrm{max\_dur})$. Recall that $G(t)$ is the subset of scenarios still possible at time-step t. Let $\delta_{i,j} = 1 - \mathrm{frac}(X^s_{i,j}), \delta'_{i,k} = \mathrm{frac}(X^s_{i,k})$ for any $s \in G(i - \mathrm{max\_dur})$.

4. Since $A^{\mathrm{aq}}_{s,t} \in \mathbb{Z}^+ \; \forall s, t$, we can partition the scenario set $G(i - \mathrm{max\_dur})$ into two exclusive subsets S^A_i, S^B_i as described below.

140

i. $\forall s' \in S_i^A$, there exists time index $p_{s'}$ s.t. $i < p_{s'} < j$ and $X_{p_{s'},j}^{s'} \notin \mathbb{Z}^+$.

ii. $\forall s' \in S_i^B$, there exists no such time index $p_{s'}$. This means $\sum_{l=j-K}^{j} X_{l,j}^{s'} \notin \mathbb{Z}^+$ (from queue balance constraint (5.3) for time j and scenario s'). Since $A_{s,j}^{\mathrm{aq}} \in \mathbb{Z}^+ \; \forall s \in S$, it is necessary that $A_{s',j}^{\mathrm{aq}} = 0$. This implies that there exists spare capacity at time j for scenario s', $A_{s',j}^{\mathrm{cap\_rem}} = A_{s',j}^{\mathrm{cap}} - \sum_{l=j-K}^{j} X_{l,j}^{s'}$, and this spare capacity $A_{s',j}^{\mathrm{cap\_rem}} \geq 1 - \mathrm{frac}(X_{i,j}^{s'}) = \delta_{i,j}$.

5. $\forall s' \in S_i^A$, select time index $p'_{s'} = \min\limits_{i < p_{s'} < j, X_{p_{s'},j}^{s'} \notin \mathbb{Z}^+} p_{s'}$, and compute
$$\delta_{p'_{s'},j} = \mathrm{frac}(X_{p'_{s'},j}^{s'}).$$

6. Compute $\delta_{\min} = \min(\delta_{i,j}, \delta'_{i,k}, \min\limits_{s' \in S_i^A} \delta_{p'_{s'},j})$.

7. Perturb solution X to X_new as follows. For all $s \in G(i - \mathrm{max\_dur}), s' \in S_i^A$,

$$
\begin{aligned}
X\_\mathrm{new}_{i,j}^{s} &= X_{i,j}^{s} + \delta_{\min} \\
X\_\mathrm{new}_{i,k}^{s} &= X_{i,k}^{s} - \delta_{\min} \\
X\_\mathrm{new}_{p'_{s'},j}^{s'} &= X_{p'_{s'},j}^{s'} - \delta_{\min} \\
X\_\mathrm{new}_{p'_{s'},k}^{s'} &= X_{p'_{s'},k}^{s'} + \delta_{\min}
\end{aligned}
\tag{5.5}
$$

Feasibility of perturbed solution: We verify the feasibility of perturbed solution X_new.

Case i. $\forall s' \in S_i^A$: The perturbation was a balanced swap of $\delta_{\min}$ units of flow between time indices j and k, and therefore no change was made to the reallocated arrival demand $\sum_{l=t-K}^{t} X_{l,t}^{s'}$ in queue balance constraint (5.3), and airborne queue $A_{s',t}^{\mathrm{aq}}$ $\forall t \in \{1, .., T\}$. Hence, the perturbed solution is feasible.

Case ii. $\forall s' \in S_i^B$: The perturbation transferred $\delta_{\min}$ units of flow from time index k to j (Recall $j < k$). However, since there was spare capacity at time index j for all scenarios $s' \in S_i^B = A_{s',j}^{\mathrm{cap\_rem}} \geq \delta_{i,j}$, a flow transfer of $\delta_{\min}$ would be fully absorbed by this spare capacity and result in no incremental airborne queue. In addition, the queue lengths for time steps k and beyond ($A_{s',t}^{\mathrm{aq}} \; \forall t \geq k$) might be reduced through

this transfer. In any case, the perturbed solution is feasible.

Objective function value for perturbed solution: We recognize that the total airborne delay cost component of objective function (5.1) for perturbed solution $X\_$new cannot be higher than that for original solution X. Therefore, the incremental ground delay cost for $X\_$new compared to X (termed $Z_g(X\_\text{new}) - Z_g(X)$) represents an upper bound in terms of the total incremental cost $Z_g(X\_\text{new}) - Z_g(X)$. The incremental ground delay cost for $X\_$new, namely, $Z_g(X\_\text{new}) - Z_g(X)$, is given by:

$$(C_{g,j-i} - C_{g,k-i} + C_{g,k-p} - C_{g,j-p})\delta_{min}, \quad \forall s' \in S_i^A \tag{5.6}$$

$$(C_{g,j-i} - C_{g,k-i})\delta_{min}, \qquad \forall s' \in S_i^B \tag{5.7}$$

$$0, \qquad \forall s' \notin G(i - \text{max\_dur}) \tag{5.8}$$

Since we have marginally non-decreasing ground-delay costs, and time indices i, p, j, k such that $i < p \le j < k$,

$$C_{g,k-i} - C_{g,j-i} \ge C_{g,k-p} - C_{g,j-p} \quad \ge \quad 0 \tag{5.9}$$

$$\implies Z_g(X\_\text{new}) - Z_g(X) \quad \le \quad 0 \tag{5.10}$$

$$\implies Z(X\_\text{new}) - Z(X) \quad \le \quad 0, \tag{5.11}$$

implying that X_new preserves optimality.

8. Adopting the above perturbation principle, Steps 1-7 can be repeated until there are no fractional values in the optimal solution. This algorithm will eventually terminate, since after every perturbation, no new fractional solution is created among $X_{a,b}^s \ \forall a, b \in \{1, .., T\}, \forall s \in S$, and at least one fractional solution among $X_{a,b}^s \ \forall a, b \in \{1, .., T\}, \forall s \in S$ is rounded off.

Lemma 1 *The hybrid stochastic formulation yields an optimal solution with integer values for all variables $X_{a,b}^s$ ($\forall s \in S$; $a, b \in \{1, .., T\}$) when the queue length variables ($A_{s,t}^{\text{aq}} \ \forall s \in S, t \in \{1, .., T\}$) are constrained to have integer values, and the ground-holding costs are marginally non-decreasing.*

142

In the original integer formulation of the hybrid stochastic model, the number of integer variables was:

Ground-holding allocation, $X_{i,j}^s \ \forall i,j \in \{1,..,T\}, s \in S$: $O(T \times T \times |S|) = O(T^3)$

Airborne queue lengths, $A_{s,t}^{\text{aq}} \ \forall t \in \{1,..,T\}, s \in S$: $O(T \times |S|) = O(T^2)$.

The total number of variables in the original formulation that need to be integral is $O(T^3)$. Lemma 1 proves that total integrality can be guaranteed by restricting the integrality requirement to $O(T^2)$ variables in the formulation.

5.4.2 Case 2: Capacity scenario tree with special structure

We consider a capacity scenario tree with a special structure, namely, one that comprises of sequentially non-decreasing capacity scenarios, with the sole element of uncertainty being the time instance when a capacity scenario branches from lowest capacity state to a higher capacity state. The scenario tree in Figure 5-2 with three capacity states (i.e., low (L), medium (M) and high (H), L<M<H) illustrates this template.
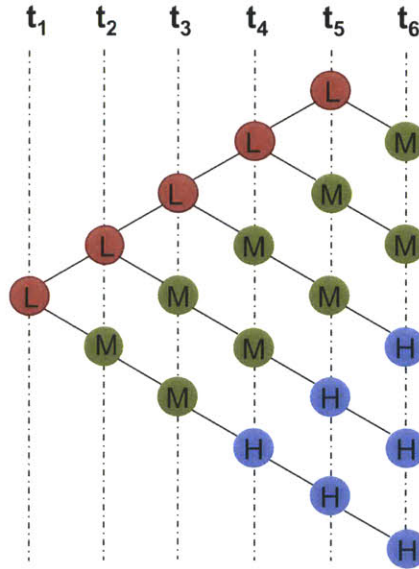
Figure 5-2: A sketch of a capacity scenario tree conforming to the special structure.

Note that every capacity scenario in the above tree follows the same deterministic trend once the capacity transitions from state (L) to state (M). That is, regardless of the time

step when capacity first increases from low state (L) to medium state (M), there are two successive medium capacity states (M) before the capacity rises to the high state (H). In the discussion that follows, we use the notation $\mathrm{dur}_s \; \forall s \in S$ to represent the duration for which lowest capacity (L) lasts for given scenario s.

Given the above template for the scenario tree, without loss of generality, we can label the scenarios numerically in the increasing order of dur_s, i.e., $S = 1, .., |S|$ where $|S| = T$ and $\mathrm{dur}_s = s \; \forall s \in \{1, .., |S|\}$. Under this notation, the lowest capacity state lasts through the first time-step for scenario 1 and the entire length of the GDP planning horizon (i.e, T intervals) for scenario T.

Now, the statement of our result is as follows:

Lemma 2 *Given*

1. *Marginally non-decreasing ground-holding cost coefficients*

   $C_{g,n+1} - C_{g,n} \geq C_{g,n} - C_{g,n-1}, \; \forall n, \; and$

2. *Capacity scenario tree with sequentially non-decreasing capacity scenarios, and sole element of uncertainty being time of improvement from lowest capacity state,*

the hybrid stochastic formulation is guaranteed to have an integral optimum solution if queue length variables for scenario T (i.e., $A^{\mathrm{aq}}_{T,t} \; \forall t \in \{1, .., T\}$) are constrained to have integer values.

5.4.2.1 Proof of Lemma 2

We provide a two-part proof of Lemma 2.

Part 1: We prove that, given specified conditions on the input parameters, the optimal solution has a special structure in terms of flight ordering with respect to the original schedule.

Part 2: We prove that, for the given special structure of optimal solution, the stated result on integrality of the optimal solution holds true.

We first develop some key notation and highlight specific properties of the hybrid formulation under the conditions mentioned in Lemma 2, that will be useful for the proof.

Notation

144

max_dur : Longest duration among all flights handled in the model.

$A_{s,b}^{\mathrm{cap\_rem}}(X)$: Residual capacity for scenario s at time-step b given solution X

$A_{s,b}^{\mathrm{cap\_rem},a}(X)$: Residual capacity for scenario s at time-step b given solution X if only flows $X_{i,j}^{s} \; \forall i \leq a - 1, \forall j$ were considered instead of the complete set $X_{i,j}^{s} \; \forall i, \forall j$.

Note: $A_{s,b}^{\mathrm{cap\_rem},T+1}(X) = A_{s,b}^{\mathrm{cap\_rem}}(X)$.

Properties of capacity scenario tree: Given the proposed scenario labeling scheme,

(A1) $G(s) = \{s, s+1, .., |S|\} \; \forall s \in \{1, .., |S|\}$ and $G(s) \backslash G(s+1) = \{s\}$ is a solitary scenario.

(A2) $A_{s,t}^{\mathrm{cap}} \geq A_{s+1,t}^{\mathrm{cap}} \; \forall s \in \{1, .., |S|\}, \; \forall t \in \{1, .., T\}$.
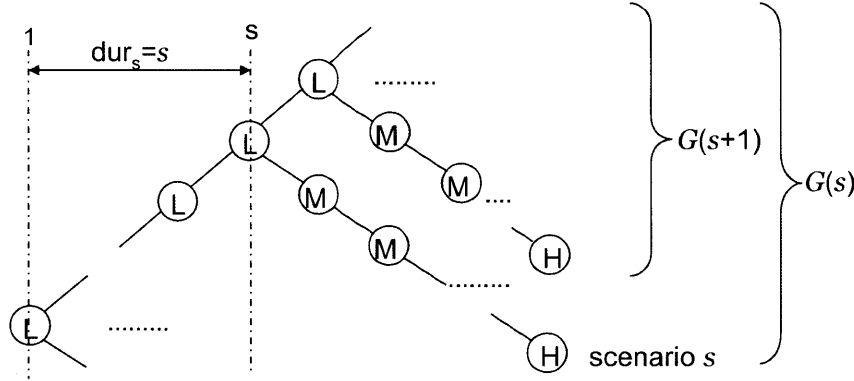
Figure 5-3: Illustration of relationship between G(s) and G(s+1) for given scenario tree.

Compact representation for ground-holding solution given scenario tree structure: The special scenario tree structure enables a compact representation of the hybrid ground-holding solution. We denote a partial solution $X^{s}(i : j)$ as the ground-holding allocation under scenario s for flights scheduled to arrive between time-steps i and j. As per the principle of the hybrid model, arrivals scheduled for time-step $t + \mathrm{max\_dur}$ are allocated ground delays at time-step t based on observed capacity values up to t. Accounting for the relevant non-anticipativity constraints, we can represent the ground-holding solution for a given scenario s as the union of two partial solutions: $X^{s}(1 : T) = [X^{T}(1 : s+\mathrm{max\_dur}) \; X^{s}(s+\mathrm{max\_dur}+1 : T)]$. In this expression, the component $X^{T}(1 : s+\mathrm{max\_dur})$ captures the portion of ground-holding decisions taken upto time-step s and is common to all scenarios indistinguishable until this time-step (i.e., $G(s)$). The subsequent decisions from

time-step $s + 1$ onwards $(X^s(s + \text{max\_dur} + 1 : T))$ are taken independently for scenario s following its divergence from scenario cluster $G(s)$.

To further assist in the proof development, we introduce the following two sub-problems at time-step t that determine partial ground-holding solutions $X^s(t + \text{max\_dur} : T)$ $\forall s \in G(t)$ and $X^{t-1}(t + \text{max\_dur} : T)$, given partial solution $X^T(1 : t + \text{max\_dur} - 1)$. Note that scenario labeled $t - 1$ corresponds to $G(t-1) \backslash G(t)$.

Figure 5-4 illustrates the two sub-problems at time t: The first (shown in green) corresponds to a deterministic ground-holding problem that is solved for the branch of the scenario tree that becomes certain at time t, while the second corresponds to the portion of the scenario tree that is still uncertain at time t.
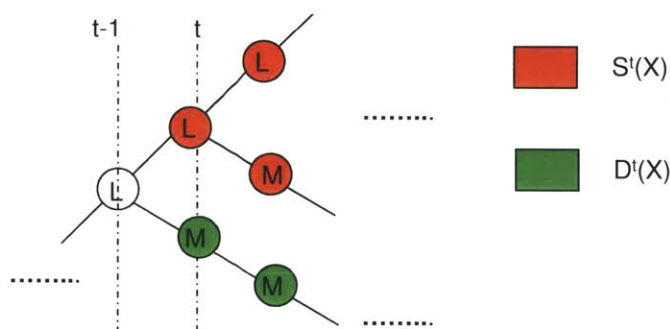
Figure 5-4: Illustration of the two sub-problems at time t for given scenario tree.

146

Deterministic ground-holding sub-problem (D_sub^t(X$_0$)): Given a partial solution $X_0^T(1 : t + \text{max\_dur} - 1)$, the partial solution $X_0^{t-1}(t + \text{max\_dur} : T)$ is the solution to the following deterministic ground-holding problem:

$$\text{Minimize} \quad \sum_{n=0}^{K} C_{g,n} \left(\sum_{i=t+\text{max\_dur}}^{T-n} X_{i,i+n}^s \right) + C_a \sum_{i=t+\text{max\_dur}}^{T} A_i^{\text{aq,D\_sub}^t} \tag{5.12}$$

subject to:

$$\sum_{j=i}^{i+K} X_{i,j}^s = A_i^d, \ \forall i \in \{t + \text{max\_dur}, .., T\} \tag{5.13}$$

$$A_i^{\text{aq,D\_sub}^t} \geq \sum_{j=\max(t+\text{max\_dur},i-K)}^{i} X_{j,i}^s + A_{i-1}^{\text{aq,D\_sub}^t}$$

$$- A_{s,i}^{\text{cap\_rem},t+\text{max\_dur}}(X_0), \ \forall i \in \{t + \text{max\_dur}, .., T\} \tag{5.14}$$

$$X_{i,j}^s \in \mathbb{Z}^+, \ \forall i, j \in \{t + \text{max\_dur}, .., T\} \tag{5.15}$$

Note that, due to the inherent property of deterministic ground-holding, flight ordering in partial solution $X_0^{t-1}(t + \text{max\_dur} : T)$ for scenario $t - 1$ in any solution X_0 will always be same as in original schedule. We denote the arrival queue length in this sub-problem for time interval i as $A_i^{\text{aq,D\_sub}^t}$.

Stochastic ground-holding sub-problem ($\mathbf{S\_sub^t(X_0)}$): Given a partial solution $X_0^T(1 : t + \text{max\_dur} - 1)$, the partial solutions $X_0^s(t + \text{max\_dur} : T) \; \forall s \in G(t)$ can be given as the solution for the following stochastic ground-holding problem:

$$\text{Minimize} \quad \sum_{s \in G(t)} \frac{p_s}{\sum_{s \in G(t)} p_s} \left(\sum_{n=0}^{K} C_{g,n} \sum_{i=t+\text{max\_dur}}^{T-n} X_{i,i+n}^s + C_a \sum_{i=t+\text{max\_dur}}^{T} A_{s,i}^{\text{aq,S\_sub}^t} \right) \quad (5.16)$$

subject to:

$$\sum_{j=i}^{i+K} X_{i,j}^s = A_i^d, \; \forall i \in \{t + \text{max\_dur}, .., T\}, s \in G(t) \quad (5.17)$$

$$A_{s,i}^{\text{aq,S\_sub}^t} \geq \sum_{j=\max(t+\text{max\_dur},i-K)}^{i} X_{j,i}^s + A_{s,i-1}^{\text{aq,S\_sub}^t} - A_{s,i}^{\text{cap\_rem},t+\text{max\_dur}}(X_0),$$

$$\forall i \in \{t + \text{max\_dur}, .., T\}, \quad s \in G(t) \quad (5.18)$$

$$X_{i,j}^{s^1} = X_{i,j}^{s^2}, \; \forall s_1, s_2 \in G(i - \text{max\_dur}), \forall i \in \{t + \text{max\_dur}, .., T\} \quad (5.19)$$

$$X_{i,j}^s \in \mathbb{Z}^+, \; \forall i, j \in \{t + \text{max\_dur}, .., T\} \quad (5.20)$$

We denote the arrival queue length in this sub-problem for time interval i by $A_{s,i}^{\text{aq,S\_sub}^t} \; \forall s \in G(t)$.

5.4.2.2 Proof of Lemma 2, Part 1

We prove the claim that there exists an optimal solution X for the hybrid formulation such that $X^T(1:T)$, the ground-holding allocation for scenario T (longest duration of lowest capacity state), has the same ordering of flights as in original schedule. The proof is based on perturbation analysis.

1. **Non-conforming solution:** Assume an optimal solution X to the hybrid model for which $X^T(1:T)$ violates the flight ordering in the original schedule. Then, there exist time intervals $i < l < k$ s.t. $X_{i,k}^T, X_{i+1,l}^T > 0$. Let's consider the lowest index i amongst all such schedule rearrangements, and the lowest corresponding indices k and l for this i. Note that ground-hold decisions for arrivals originally scheduled at i are made at $i - \text{max\_dur}$.

   Now, we know that scenario labeled $i - \text{max\_dur} = G(i - \text{max\_dur}) \backslash G(i+1 - \text{max\_dur})$ is a solitary scenario (from (A1)) and $A_{i-\text{max\_dur},t}^{\text{cap}} \geq A_{s,t}^{\text{cap}} \ \forall s \in G(i - \text{max\_dur}), t \in \{1,..,T\}$ (from (A2)). For scenario $i - \text{max\_dur}$ and solution X, only one of the following four cases is possible.

   Case (a): $A_{i-\text{max\_dur},l}^{\text{cap\_rem}}(X) > 0$ that can accommodate additional δ units of flow without producing any queue.

   Case (b): $A_{i-\text{max\_dur},l}^{\text{cap\_rem}}(X) = 0$, but $\exists p \geq i + 1, p \leq l$ s.t. $X_{p,l}^{i-\text{max\_dur}} > 0$.

   Case (c): $X_{p,l}^{i-\text{max\_dur}} = 0 \ \forall p \geq i + 1$ and $A_{i-\text{max\_dur},l}^{\text{cap\_rem}}(X) = 0$, but $A_{i-\text{max\_dur},l}^{\text{cap\_rem},i+1}(X) > 0$.

   Case (d): $X_{p,l}^{i-\text{max\_dur}} = 0 \ \forall p \geq i + 1$ and $A_{i-\text{max\_dur},l}^{\text{cap\_rem}}(X) = 0$ and $A_{i-\text{max\_dur},l}^{\text{cap\_rem},i+1}(X) = 0$.

2. **Perturbation:** Consider a new solution X_new obtained by swapping infinitisemal (δ) units of ground-hold allocation between $X_{i,k}^T$ and $X_{i+1,l}^T$, as described below.

$$X\_\text{new}_{i,k}^T = X_{i,k}^T - \delta, \quad X\_\text{new}_{i,l}^T = X_{i,l}^T + \delta$$

$$X\_\text{new}_{i+1,k}^T = X_{i+1,k}^T + \delta, \quad X\_\text{new}_{i+1,l}^T = X_{i+1,l}^T - \delta$$

149

Note that this would amount to a balanced swap for all scenarios except $i - \text{max\_dur}$. For case (b) alone, consider an additional perturbation involving a balancing transfer of δ units from $X_{p,l}^{i-\text{max\_dur}}$ to $X_{p,k}^{i-\text{max\_dur}}$ for scenario $i - \text{max\_dur}$.

$$X\_\text{new}_{p,l}^{i-\text{max\_dur}} = X_{p,l}^{i-\text{max\_dur}} - \delta$$
$$X\_\text{new}_{p,k}^{i-\text{max\_dur}} = X_{p,k}^{i-\text{max\_dur}} + \delta$$

3. **Feasibility of perturbation:** Since the perturbation is essentially a rearrangement of ground-holding allocation for arrivals scheduled for time indices i and $i + 1$ (and p in case (b)), its feasibility is not affected in any way.

4. **Cost of Perturbation:** Note that the perturbations only affect scenario set $G(i - \text{max\_dur})$, as it features changes to arrivals scheduled at and beyond time index i for scenario T. We compute the net cost of perturbation specific to each scenario within this affected set across each of the above-described cases (a)-(d).

1. $\forall s \in G(i + 1 - \text{max\_dur})$:

The perturbation is effectively a balanced swap of ground-hold allocations (between $X_{i,k}^T$ and $X_{i+1,l}^T$). Therefore, the airborne queue length variables ($A_{s,t}^{\text{aq}}, \forall s \in G(i + 1 - \text{max\_dur})$) are not altered, and the net cost of this perturbation is restricted to ground-hold cost coefficients.

Cost of perturbation = $(C_{g,l-i} - C_{g,k-i} + C_{g,k-(i+1)} - C_{g,l-(i+1)})\delta \le 0$ given marginally non-decreasing ground-delay costs and $k > l \ge i + 1$.

2. For $s = i - \text{max\_dur}$:

Case a): The unbalanced δ units of ground-hold re-allocation to time index l (from $X_{i,k}^T$) are absorbed by the available spare capacity (since $A_{i-\text{max\_dur},l}^{\text{cap\_rem}}(X) > 0$) without producing any queue. Therefore, the perturbation causes no increase to airborne delay costs.

Highest cost of perturbation = $(C_{g,l-i} - C_{g,k-i})\delta < 0$ (corresponds to situation where the perturbation causes no decrease to airborne delay costs (i.e. $A_{i-\text{max_dur},k}^{\text{aq}}(X) = $

0)).

Case b): The perturbation is effectively a balanced swap of ground-hold allocations (between $X_{i,k}^T$ and $X_{i+1,l}^{i-\text{max\_dur}}$). Recall the compact representation for ground-holding solution $X^{i-\text{max\_dur}}(1:T) = [X^T(1:i) \; X^{i-\text{max\_dur}}(i+1:T)]$. Therefore, the airborne queue length variables ($A_{i-\text{max\_dur},t}^{\text{aq}}$) are not altered, and the net cost of this perturbation is restricted to ground-hold cost coefficients.

Cost of perturbation $= (C_{g,l-i} - C_{g,k-i} + C_{g,k-p} - C_{g,l-p})\delta \leq 0$ for marginally non-decreasing ground-delay costs.

Case c): Since there is no available spare capacity at time l ($A_{i-\text{max\_dur},l}^{\text{cap\_rem}}(X) = 0$), we know the unbalanced δ units of ground-hold re-allocation will produce an incremental queue (i.e. $A_{i-\text{max\_dur},l}^{\text{aq}}$ will increase by δ). In the worst-case, when there is no spare capacity at any of time indices $l \leq t \leq k$ for scenario $i - \text{max\_dur}$, the incremental queue will last through to time k (i.e. $A_{i-\text{max\_dur},t}^{\text{aq}} \forall l \leq t \leq k$ will all increase by δ).

Therefore, highest cost of perturbation $= (C_{g,l-i} - C_{g,k-i} + C_a(k - l))\delta$.

We will now show that this worst-case cost of perturbation is non-positive.

From our description of the deterministic sub-problem D_sub^t, we know that $X^{i-\text{max\_dur}}(i+1:T)$ can be obtained as solution to sub-problem $\text{D\_sub}^{i+1-\text{max\_dur}}(X)$. For this case, it is also known that $A_{i-\text{max\_dur},l}^{\text{cap\_rem},i+1}(X) > 0$ (there is spare capacity at l if we ignore all arrivals originally scheduled beyond i). Since for this case, $A_{i-\text{max\_dur},l}^{\text{cap\_rem}}(X) = 0$ and $X_{p,l}^{i-\text{max\_dur}} = 0 \quad \forall p \geq i+1$, we can conclude that, in sub-problem $\text{D\_sub}^{i+1-\text{max\_dur}}(X)$, $A_{l-1}^{\text{aq,D\_sub}^{i+1-\text{max\_dur}}} > 0$.

In other words, we know there are some arrivals originally scheduled to land at time $r \geq i+1$ that are re-allocated to a time $h < l$ under scenario $i - \text{max\_dur}$ that result in incremental airborne queues and utilize the spare capacity $A_{i-\text{max\_dur},l}^{\text{cap\_rem},i+1}(X)$. Lets denote the latest such time instance by h, and latest corresponding scheduled arrival time by r.

Let us consider an alternate feasible solution X_alt to the sub-problem $\text{D\_sub}^{i+1-\text{max\_dur}}(X)$

151

where a small portion α of the arrivals scheduled for time r ($\alpha \leq A^{\text{cap\_rem},i+1}_{i-\text{max\_dur},l}(X)$) are now reallotted, through ground-hold, to time l instead of h.

$$X\_alt^{i-\text{max\_dur}}_{r,h} = X^{i-\text{max\_dur}}_{r,h} - \alpha \qquad (5.21)$$

$$X\_alt^{i-\text{max\_dur}}_{r,l} = X^{i-\text{max\_dur}}_{r,l} + \alpha, \qquad (5.22)$$

all else being equal. In this alternate solution, the airborne delays from time-step h to l for the reallotted α units are replaced by ground-delays. Therefore, change in objective function value for D_sub$^{i+1-\text{max\_dur}}(X)$ is given by Z(X_alt)-Z(X)= $(C_{g,l-r} - C_{g,h-r} - C_a(l-h))\alpha$.

Since the original solution $X^{i-\text{max\_dur}}(i+1:T)$ was optimal for the sub-problem D_sub$^{i+1-\text{max\_dur}}(X)$, Z(X_alt)-Z(X) ≥ 0, that is, $C_a(l-h) \leq (C_{g,l-r} - C_{g,h-r})$. Given marginally non-decreasing ground delay costs and $r > i$, we have $C_a(l-h) \leq C_{g,l-r} - C_{g,h-r} \leq C_{g,l-i} - C_{g,h-i}$.

Also, given $h < l < k$, we have $C_a(k-l) \leq C_{g,k-i} - C_{g,l-i}$.

Therefore, the highest cost of perturbation = $(C_{g,l-i} - C_{g,k-i} + C_a(k-l))\delta \leq 0$.

Case d): The nature of perturbation is the same as in case (c). Therefore, the highest cost of perturbation = $(C_{g,l-i} - C_{g,k-i} + C_a(k-l))\delta$. We need to once again show that this worst-case cost of perturbation is non-positive.

From our description of the stochastic sub-problem S_sub^t, we know that $X^s(i+1:T)$, $\forall s \in G(i+1-\text{max\_dur})$ can be obtained as solution to sub-problem S_sub$^{i+1-\text{max\_dur}}(X)$.

For this case, it is given that $A^{\text{cap\_rem},i+1}_{i-\text{max\_dur},l}(X) = 0$. Coupling this with property (A2) of the scenario tree, we can conclude that $A^{\text{cap\_rem},i+1}_{s,l}(X) = 0$ $\forall s \in G(i-\text{max\_dur})$. We are also given that $X^T_{i+1,l} > 0$. This implies that, in sub-problem S_sub$^{i+1-\text{max\_dur}}(X)$, $A^{\text{aq,S\_sub}^{i+1-\text{max\_dur}}}_{s,l} > 0$ $\forall s \in G(i+1-\text{max\_dur})$ (since there is no spare capacity at time step l for any scenario in set $G(i+1-\text{max\_dur})$ even before servicing arrivals scheduled beyond time i).

Consider an alternate feasible solution $X\_alt^s(i+1:T)$ to sub-problem S_sub$^{i+1-\text{max\_dur}}(X)$

152

obtained by reallotting, through ground-holds, a portion α of the arrivals scheduled for time $i + 1$ to time $l + 1$ instead of l for scenario T. That is, consider

$$X\_alt^T_{i+1,l+1} = X^T_{i+1,l+1} + \alpha \qquad (5.23)$$

$$X\_alt^T_{i+1,l} = X^T_{i+1,l} - \alpha, \qquad (5.24)$$

all else being equal. For this alternate solution, the airborne delays from time step l to $l + 1$ for the reallotted α units are replaced by ground-delays for all scenarios $s \in G(i + 1 - \text{max\_dur})$. Therefore, the change in objective function value for S_sub$^{i+1-\text{max\_dur}}(X)$: Z(X_alt) - Z(X) = $(C_{g,l+1-(i+1)} - C_{g,l-(i+1)} - C_a)\alpha$.
Since $X^s(i+1 : T)$ $\forall s \in G(i+1-\text{max\_dur})$ is an optimal solution to S_sub$^{i+1-\text{max\_dur}}(X)$, we know that

Z(X_alt)$\geq$Z(X) $\Rightarrow C_{g,l-i} - C_{g,l-i-1} \geq C_a$.

Given $k > l > i$ and marginally non-decreasing ground delay costs, we have

$C_{g,k-i} - C_{g,l-i} \geq C_a(k - l)$.

Therefore, the highest cost of perturbation,

$(C_{g,l-i} - C_{g,k-i} + C_a(k - l))\delta \leq 0$.

The above cases prove that Z(X_new)$\leq$Z(X). We can repeat the above-described form of perturbation till we have an optimal solution such that the ground-holding allocation for scenario T (longest duration of lowest capacity state), $X^T(1 : T)$, has the same ordering of flights as in original schedule.

5.4.2.3 Proof of Lemma 2, Part 2

We now have to prove that an optimal solution X is integral under the additional condition that $A^{\text{aq}}_{T,t} \in \mathbb{Z}^+ \ \forall t \in \{1, .., T\}$.

The proof in Part 1 concerning the structure for the optimal solution X holds for any general value of $A^{\text{aq}}_{s,t} \ \forall s \ \forall t$. Therefore, the structure holds true for specific case of integral $A^{\text{aq}}_{T,t}, \ \forall t$.

We again adopt a perturbation analysis for this proof.

1. **Non-conforming solution:** Assume we have a non-integral optimal solution X such that for scenario T (longest duration of lowest capacity state) there exist time instances $p \in \{1, ..T\}$, $j \in \{p, p+1, .. \min(p+K, T)\}$ such that $X_{p,j}^T \notin \mathbb{Z}^+$. Let us select the earliest such time instance p, and corresponding earliest time instance j for which $X_{p,j}^T \notin \mathbb{Z}^+$. In accordance to the structure for X as derived in Part 1, the ordering for flights in X for scenario T is the same as in original schedule.

Since, for a given p, j is the lowest time index for which $X_{p,j}^T \notin \mathbb{Z}^+$, $\exists$ time instance $q > j$ s.t. $X_{p,q}^T > 0$, since $\displaystyle\sum_{t=p}^{\min(p+K,T)} X_{p,t}^T = A_p^d$ where $A_p^d \in \mathbb{Z}^+$. Let q be the lowest such time instance.

Given the order preserving structure of the solution for scenario T, we can infer that no arrival originally scheduled beyond time p is allotted to any time at or before j. i.e. $X_{k,t}^T = 0 \ \forall k > p, t \le j$.

Also, since p is the lowest time index for which $X_{p,j}^T \notin \mathbb{Z}^+$, we have $X_{k,l}^T \in \mathbb{Z}^+ \ \forall k < p, l \in \{k, k+1, .. \min(k+K, j)\}$. Therefore,

$$\sum_{t=\max(1,j-K)}^{j} (X_{t,j}^T) = \sum_{t=\max(1,j-K)}^{p-1} (X_{t,j}^T) + X_{p,j}^T \notin \mathbb{Z}^+, \text{ since } \sum_{t=\max(1,j-K)}^{p-1} (X_{t,j}^T) \in \mathbb{Z}^+.$$

We know that $A_{T,j}^{\mathrm{aq}} = \min(0, \displaystyle\sum_{t=\max(1,j-K)}^{j} X_{t,j}^T + A_{T,j-1}^{\mathrm{aq}} - A_{T,j}^{\mathrm{cap}})$,

where $A_{T,j}^{\mathrm{cap}}$ and $A_{T,j-1}^{\mathrm{aq}} \in \mathbb{Z}^+$.

Therefore, if $A_{T,j}^{\mathrm{aq}} \in \mathbb{Z}^+$, the only possibility is that $A_{T,j}^{\mathrm{aq}} = 0$. This implies that capacity is not exceeded at time j for scenario T, that is,

$$A_{T,j}^{\mathrm{cap\_rem}, p+1}(X) = A_{T,j}^{\mathrm{cap}} - A_{T,j-1}^{\mathrm{aq}} - \sum_{t=\max(1,j-K)}^{j} (X_{t,j}^T) \ge 0 \text{ and } \notin \mathbb{Z}^+.$$

The above, in turn, implies $A_{T,j}^{\mathrm{cap\_rem}, p+1}(X) > 0$.

From property (A2) of scenario tree, we can conclude $A_{s,j}^{\mathrm{cap\_rem}, p+1} > 0 \ \forall s \in \{1, .., T\}$.

As per the hybrid stochastic model's working principle, the ground-holding decision $X_{p,j}^T$ is taken at time $p - \mathrm{max\_dur}$, and affects scenarios in set $G(p - \mathrm{max\_dur})$. There are three possible categories of scenarios within $G(p - \mathrm{max\_dur})$:

154

Type 1. $s \in G(p - \text{max\_dur})$ such that $A_{s,j}^{\text{cap-rem}}(X) > 0$

Type 2. $s \in G(p - \text{max\_dur})$ such that $A_{s,j}^{\text{cap-rem}}(X) = 0$, but $\exists m$ such that $p \leq m < j$ and $X_{m,j}^s > 0$

Type 3. $s \in G(p - \text{max\_dur})$ such that $A_{s,j}^{\text{cap-rem}}(X) = 0$, and $\nexists m$ such that $p \leq m < j$ and $X_{m,j}^s > 0$

Note that scenario T falls into Type 1.

2. **Perturbation:** Consider a new solution X_new obtained by advancing δ units of ground-hold allocation $X_{p,q}^T$ to $X_{p,j}^T$ in the following manner.

$$X\_\text{new}_{p,j}^T = X_{p,j}^T + \delta \tag{5.25}$$

$$X\_\text{new}_{p,q}^T = X_{p,q}^T - \delta, \tag{5.26}$$

all else being equal. For scenario $s \in G(p - \text{max\_dur})$ belonging to Type 2, we consider additional perturbation involving a balancing transfer of δ units from $X_{m,j}^s$ to $X_{m,q}^s$.

$$X\_\text{new}_{m,j}^s = X_{m,j}^s - \delta \tag{5.27}$$

$$X\_\text{new}_{m,q}^s = X_{m,q}^s + \delta, \tag{5.28}$$

all else being equal.

3. **Feasibility of perturbation:** Given the perturbation is essentially a rearrangement of ground-holding allocation for arrivals scheduled for time indices p under scenario T (and m for scenario s belonging to Type 2), its feasibility is not affected in any way.

4. **Cost of perturbation:** We now consider cost of perturbation specific to scenarios from each of the above three categories.

   Type 1 Identical reasoning to case (a) from proof in Part 1. We can thereby show highest cost of perturbation $= Z_s(X\_\text{new}) - Z_s(X) = (C_{g,j-p} - C_{g,q-p})\delta < 0$.

   Type 2 Identical reasoning to case (b) from proof in Part 1. We can thereby show cost

of perturbation $= Z_s(X\_\text{new}) - Z_s(X) = (C_{g,j-p} - C_{g,q-p} + C_{g,q-m} - C_{g,j-m})\delta \leq 0$ for marginally non-decreasing ground delay costs.

Type 3 Identical reasoning to case (c) from proof in Part 1. We can thereby show $Z_s(X\_\text{new}) - Z_s(X) = (C_{g,j-p} - C_{g,q-p} + C_a(q - j))\delta \leq 0$.

Therefore, $Z(X\_\text{new}) \leq Z(X)$.

We can repeat the above-described form of perturbation till we have an optimal solution that bears only integral values for $X_{p,j}^T \ \forall p \in \{1, .., T\}, j \in \{p, p + 1, .., \min(p + K, T)\}$.

As shown earlier, the compact representation for ground-holding solution for any scenario $s \in \{1, .., |S|\}$ is $X^s(1 : T) = [X^T(1 : s + \text{max\_dur}) \quad X^s(s + \text{max\_dur} + 1 : T)]$, where the partial solution $X^s(s + \text{max\_dur} + 1 : T)$ can be obtained as solution to the deterministic sub-problem $D\_\text{sub}^{s+1}(X)$.

Given integral values for $X^T(1 : T)$, we know that $A_{s,t}^{\text{cap\_rem},s+1+\text{max\_dur}}(X) \in \mathbb{Z}^+ \ \forall t \geq s + 1 + \text{max\_dur}$. Therefore, the solution to the deterministic sub-problem $D\_\text{sub}^{s+1}(X)$ will also be integral for all s, ensuring that the overall ground-holding solution $X^s(1 : T)$ will be integral for all s.

In summary, we have shown using Parts 1 and 2 of this proof that, if the queue length variables for scenario T (longest duration of lowest capacity state) are restricted to be integral, the hybrid stochastic ground-holding model will yield an integral optimum under

1. Marginally non-decreasing ground-holding cost coefficients, and

2. Capacity scenario tree with sequentially non-decreasing capacity scenarios, with the sole element of uncertainty being the time of improvement from the lowest capacity state.

Therefore, total integrality under these conditions can be guaranteed by restricting the integrality requirement to $O(T)$ variables in the formulation, instead of $O(T^3)$ variables in the original formulation.

A key motivation for developing the hybrid stochastic model was the notional tradeoff in pre-CDM vs. CDM performances of the static and dynamic models, which the hybrid model seeks to balance. We conjecture that the hybrid stochastic model inherits some of the

favorable properties of both static and dynamic models, and may therefore produce better post-CDM performance. The next chapter studies this conjecture and the factors influencing it in greater detail.

Chapter 6

Application of CDM to Stochastic Ground-Holding Models

6.1 Introduction

The objective of this chapter is to study the application of Collaborative Decision-Making (CDM) mechanisms to the stochastic ground-holding models discussed in the previous chapter, and analyze the pros and cons of each model through representative case studies.

CDM mechanisms are integral features of a modern Ground Delay Program (GDP) that allow airlines to participate in the slot allocation process. The concept of CDM within a GDP has developed rapidly in recent years and has been unanimously acknowledged to have benefited airport capacity users [18]. Through the mechanism of intra-airline slot substitution, an airline can unilaterally redistribute slots allotted to its flights in accordance to flight-specific delay preferences. Through the mechanism of compression (and its real-time version Slot Credit Substitution (SCS)), airlines can trade slots allotted to delayed or cancelled flights in exchange for a later slot. Both CDM mechanisms help airlines further reduce their delay costs by allowing them to incorporate privately-held information on flight-specific delay costs that are not available to the airports for the ground-hold allocation process [18].

As seen in Chapter 4, *static* [67] and *dynamic* [55] stochastic ground-holding models are the current state-of-the-art from literature, while the *hybrid stochastic model* was developed

in the Chapter 5 as a compromise between these two. These models allocate ground delays under capacity uncertainty with the objective of minimizing expected total delay costs assuming homogeneous unit delay costs for all flights. The key differences in the solution properties of the three models and the consequent tradeoffs with respect to CDM mechanisms were highlighted in the previous chapter. The dynamic model, unlike the static model, is a multi-stage formulation that allocates scenario-specific ground delays, but it does so by differentiating between flights based on individual flight durations. Hence, while it guaranteed to achieve lower system delay costs given the same homogeneous unit delay costs, the dynamic solution offers limited flexibility for the operating airlines to swap slots between flights of differing durations. The hybrid model was developed with the purpose of bridging this tradeoff by combining the pre-CDM efficiency of the dynamic model and CDM-flexibility of the static model.

This chapter looks to assess the theorized advantages of the hybrid model by comparing its performance with the static and the dynamic models when applied to GDP case studies spanning a range of input data. In the currently practiced GDP framework, the CDM mechanisms are designed for deterministic settings. In order to construct a complete GDP case study involving stochastic ground-holding models, variants of CDM mechanisms compatible with static, dynamic and hybrid stochastic ground-holding solutions are developed. Appropriate metrics that capture system-wide delay benefits generated by the CDM mechanisms are used to evaluate the post-CDM performances of the three stochastic ground-holding models, and to determine their relative merits. The hybrid stochastic model is rarely observed to be the worst-performing of the three models across the range of GDP cases studied, emphasizing its ability to present a robust compromise between the static and dynamic models.

Section 6.2 of this chapter describes the proposed GDP framework involving stochastic ground-holding models, along with details of information sharing between airport and airlines regarding the stochastic ground-holding solutions. Sections 6.3 and 6.4 discuss the mechanisms of intra-airline slot substitution and SCS/Compression respectively, starting with the formulations for their stochastic versions followed by details of the associated case studies comparing the performances of three stochastic models when combined with the respective

mechanisms.

6.2   GDP Framework within the CDM Paradigm

As illustrated in Figure 6-1 below, a GDP can be viewed as an interaction between an airport and the airlines whose flights are rescheduled by the GDP. The airport enforcing a ground delay program uses inputs on scheduled arrival demand and forecasted capacity over a future time horizon to determine the controlled slot allocation through ground-holding. A Ration-By-Schedule (RBS) principle is followed in assigning slots to individual flights. Given this allocation, airlines operating the delayed flights are allowed to engage in intra-airline slot substitutions and report changes to the airport. SCS is a real-time, dynamic version of Compression, and is an event-driven process executed whenever an airline expresses the desire to forfeit an earlier slot in exchange for a later slot. The airport authorities accommodate this request by feasibly advancing flights occupying the intervening slots.
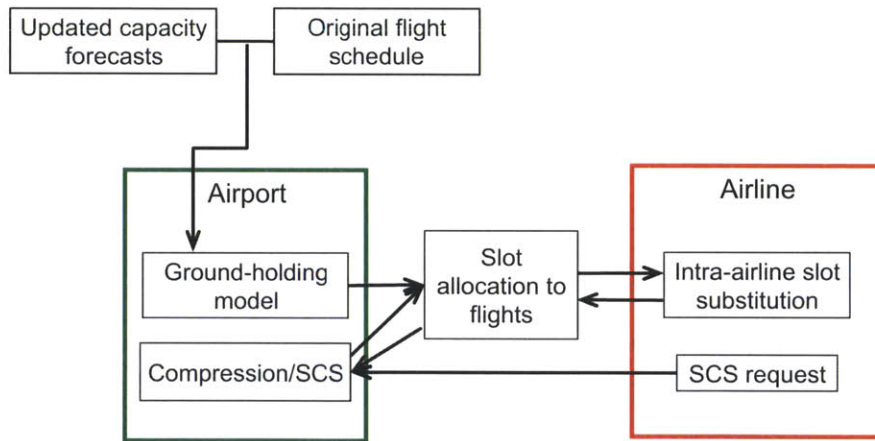
Figure 6-1: Operational framework for a GDP in the CDM paradigm.

In current GDP practice, uncertainty in capacity is handled by rerunning the ground-holding model and generating a renewed slot allocation whenever there is an update in the forecasted future capacity (which is assumed to be deterministic following every update). Any alteration to the slot allocation, triggered either by a capacity forecast update, intra-airline slot substitution or SCS event, is communicated between the airport and the operating airlines.

A few past studies featuring capacity uncertainty have adopted such an iterative framework [6, 56]. The present study, however, looks to develop a GDP framework in which stochastic ground-holding models are used for generating the initial slot allocation to which the CDM mechanisms are then applied. To the best of our knowledge, only Mukherjee and Hansen (2007) [55] have pursued this area of research and proposed stochastic equivalents for intra-airline slot substitution and compression. The formulations for the CDM mechanisms presented here are similar to those developed in Mukherjee and Hansen (2007) [55].

6.2.1 Stochastic ground-holding models in the GDP framework

As described in Chapters 4 and 5, stochastic ground-holding models allocate arrival slots while explicitly planning for uncertainty in capacity forecasts. Like their deterministic counterparts, these models assume homogeneous unit delay costs across all flights, and minimize the expected sum of ground and airborne delay costs across all capacity scenarios. They generate a slot allocation that applies to all the capacity scenarios considered in the original forecast. The use of such allocations at the start of the GDP would therefore reduce the number of allocation revisions triggered by capacity forecast updates. Indeed, if the original set of capacity scenarios and their occurrence probabilities are accurately forecasted, there would be no further need for revising the initial slot allocation except for airline responses.

This study focuses on three stochastic ground-holding models described in the previous chapter: the static [67], the dynamic [55], and the hybrid (developed in the Chapter 5). In a key simplification, the unit delay costs are assumed to be linear. This assumption reduces the volume of information pertaining to the slot allocation that needs to be exchanged between airport and airlines in the stochastic GDP framework, as described in the following section.

6.2.2 Communication of slot allocation information

In the deterministic GDP framework, the slot allocation is conveyed in terms of fixed slot timings. Stochastic ground-holding solutions however embed information regarding the underlying uncertainty. While the slot allocation from the static stochastic model has the same structure as a deterministic allocation, it in fact contains underlying information on airborne

delays that varies by scenario. The slot timings from the dynamic and hybrid models themselves vary by scenario. These aspects mean a greater amount of information needs to be communicated between the airport and airlines when stochastic ground-holding models are used for slot allocation. To keep the volume of this communication manageable, the relevant information regarding each slot slot_f allotted for flight $f \in F$ in the original schedule are condensed into the following five attributes.

(a) Earliest allotted arrival time across all scenarios: $\text{EAT}(\text{slot}_f)$

(b) Expected incremental ground delay beyond $\text{EAT}(\text{slot}_f)$ across all scenarios: $\text{grd}(\text{slot}_f)$

(c) Expected airborne delay across all scenarios (assuming first-come first-serve policy for servicing airborne queues): $\text{air}(\text{slot}_f)$

(d) Duration of assigned flight: $\text{dur}(\text{slot}_f)$

(e) Earliest time of arrival for assigned flight: $\text{ETA}(\text{slot}_f)$

Assuming linear ground and airborne delay costs for all flights, the first four attributes are sufficient information for an airline to execute intra-airline slot substitution on any of the three stochastic model allocations, as described below. The last attribute is required by the airport to determine feasible slot advancements in the SCS process, as described in Section 6.4.

6.3 Intra-Airline Slot Substitution

As mentioned earlier, the intra-airline slot substitution is a process through which each airline redistributes its flights amongst the slots allotted to it. This helps the airline advance flights for which delays are prohibitive at the expense of flights for which delays are less critical. A practical instance of such a disparity in flight delay costs could be found with an airline that operates a flight with predominantly connecting passengers along with another carrying a majority of local passengers. In this case, the airline could potentially favor delay reduction for the first flight in exchange for corresponding delay increase to the second flight.

The intra-airline slot substitution can be viewed as an assignment problem solved by each airline (labelled a) to assign flights operated by it ($F_a \subset F$) to the slots allotted to it ($\text{slot}_f \; \forall f \in F_a$), given its knowledge of flight-specific unit delay costs ($C_g^f, C_a^f, \; \forall f \in F_a$) and swapping feasibility ($\text{feas}_{f,k} \; \forall f, k \in F_a$) determined by the slot allocation.

In the context of a stochastic slot allocation, swapping feasibility $\text{feas}_{f,k}$ is governed by whether a flight f can feasibly occupy a slot slot_k's timings across all scenarios. The determination of this operational feasibility depends upon the type of stochastic ground-holding model used for obtaining the slot allocation, as is described below:

Static : $\text{feas}_{f,k} = 1$, if $\text{arr}_f \leq \text{ETA}(\text{slot}_k)$

Hybrid : $\text{feas}_{f,k} = 1$, if $\text{arr}_f \leq \text{ETA}(\text{slot}_k)$

Dynamic : $\text{feas}_{f,k} = 1$, if $\text{arr}_f \leq \text{ETA}(\text{slot}_k)$ and $\text{dur}_f = \text{dur}_k$

The definitions for the feasibility parameter $\text{feas}_{f,k}$ highlight the reduced amenability of the dynamic model for slot substitution compared to the static and hybrid models. Since the slot allocation of the dynamic model depends upon individual flight durations, two flights of different durations cannot always have their slots swapped. This phenomenon was elaborated upon in the previous chapter.

The cost of assigning a flight $f \in F_a$ to slot $\text{slot}_k, k \in F_a$, denoted by $C_{f,k}$, can be computed as follows:

$$C_{f,k} = C_g^f(\text{grd}(\text{slot}_k) + \text{ETA}(\text{slot}_k) - \text{arr}_f) + C_a^f \text{air}(\text{slot}_k) \tag{6.1}$$

The assignment formulation for the intra-airline substitution follows as described below.

$$\text{Minimize} \quad \sum_{f \in F_a} \sum_{k \in F_a} C_{f,k} X_{f,k} \quad (6.2)$$

subject to:

$$\sum_{f \in F_a} X_{f,k} = 1, \ \forall k \in F_a \quad (6.3)$$

$$\sum_{k \in F_a} X_{f,k} = 1, \ \forall f \in F_a \quad (6.4)$$

$$X_{f,k} \leq \text{feas}_{f,k}, \ \forall f, k \in F_a \quad (6.5)$$

$$X_{f,k} \in 0, 1 \ \forall f, k \in F_a$$

where $X_{f,k}$ is a binary indicator variable signifying the assignment of flight f to slot$_k$.

This assignment formulation is simpler to the one presented in Mukherjee and Hansen (2007) [55] for intra-airline slot substitution, in that airlines are not permitted to swap slots in a scenario-specific manner in the above formulation. This simplification was adopted to facilitate the compact structure for the slot allocation information described in Section 6.2.2. In order to support scenario-specific swapping of slots in intra-airline substitution, scenario-specific arrival times for each slot slot$_f$ would need to be communicated to the airlines. The above formulation does not explicitly model flight cancellation decisions alongside slot substitutions, as was done in Mukherjee and Hansen (2007) [55]. While the formulation can be easily extended to include cancellation decisions, this aspect is not analyzed in the following case studies. Lastly, given the formulation design, intra-airline substitution would only effect changes to attributes (d) and (e) of the slot allocation information described in Section 6.2.2.

For the same GDP input, the static, hybrid and dynamic models typically provide different slot allocation inputs and permit varying levels of flexibility for intra-airline slot substitution. The next two sub-sections consider different GDP cases constructed from hypothetical and real-world inputs, and compare the final delay costs realized for the three models from the combination of ground-hold allocation and intra-airline slot substitution for these cases.

6.3.1 Hypothetical case studies

The following sub-sections present the base data, experimental design as well as results of hypothetical case studies, where sample input data are generated in a controlled manner to help analyze specific trends.

6.3.1.1 Base data

The key inputs for a GDP comprise of capacity forecasts, arrival schedule and flight delay costs. The capacity forecast is represented using a scenario tree. In the hypothetical case studies, arrival capacity is assumed to take one of two values: a normal value (16), and a deteriorated value (8). All scenarios begin with capacity at the deteriorated state, and differ based on the time interval after which capacity improves to the normal state. Accordingly, in the scenario tree, a scenario diverges out from the main branch at the time interval when the capacity improves. A GDP planning horizon of 10 time intervals (with each interval measuring an arbitary time unit) is created, which produces a total of 10 possible capacity scenarios with the capacity improvement happening after the $1^{st}, 2^{nd}, ..., 10^{th}$ interval respectively. The corresponding scenario tree is depicted in Figure 6-2. A scenario is labelled as $s \in \{1, .., 10\}$, where s represents the duration of deteriorated capacity.
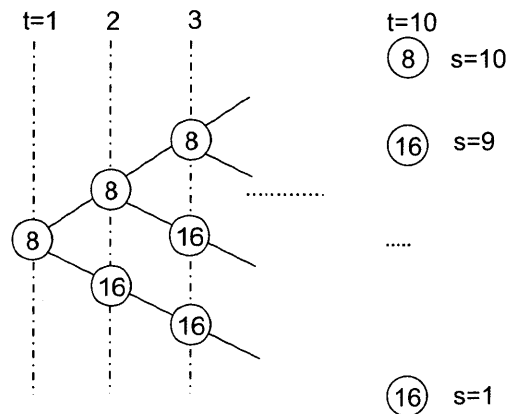
Figure 6-2: Scenario tree format for hypothetical case studies

The following arrival demand is set for the 10 time intervals: 16,16,15,16,15,16,16,15,8,5 with flight durations ranging from 2 to 5 time intervals in length. The total of 138 flights

166

are arbitrarily split across 5 different airlines. Airline 1 having a fleet of 25 arrivals out of 138 is the focus of our analysis. For simplicity of analysis, the average unit ground (C_g^f) and airborne (C_a^f) delay costs amongst flights operated by each of the 5 airlines are uniformly set at 0.5 and 2.5 respectively.

6.3.1.2 Experimental design

The objective behind these hypothetical case studies is to assess the influence of various GDP input parameters on the comparative performances of the three stochastic models. Those pertinent input parameters are:

P1 Expected duration of deteriorated capacity (as determined by the probability distribution of the capacity scenarios).

P2 Coefficient of variation in unit delay costs across flights (given mean ground delay cost=0.5; airborne delay cost=2.5).

For representing the range of the parameter P1, three GDP cases with differing probability distributions $p_s \forall s \in \{1, .., 10\}$ are generated.

Case 1: Short duration for deteriorated capacity, $p_s = \begin{cases} 0.4 & \forall s \in \{1, 2\}, \\ 0.025 & \forall s \in \{3, .., 10\} \end{cases}$

Case 2: Medium duration for deteriorated capacity, $p_s = 0.1 \quad \forall s \in \{1, .., 10\}$.

Case 3: Long duration for deteriorated capacity, $p_s = \begin{cases} 0.025 & \forall s \in \{1, 8\}, \\ 0.4 & \forall s \in \{9, 10\} \end{cases}$

First, the focus is on the parameter P1: the slot allocation results from the three stochastic ground-holding models when applied to each of the three capacity cases are determined. The focus is then centered on airline 1, and for every slot allocation obtained in the previous step, the improvement to its total delay costs through intra-airline substitution is studied, while varying parameter P2 (coefficient of variation in unit delay costs) from 0% to 30%.

6.3.1.3 Results

Before intra-airline substitution: Adopting the given mean ground and airborne unit delay costs along with the specified demand schedule, Tables 6.1, 6.2 and 6.3 present the

delay costs for the optimal slot allocations from the three stochastic models for each capacity case.

	Static	Hybrid	Dynamic
Total ground-delay costs	69	88	91
Total airborne-delay costs	86	51.5	15

Table 6.1: Static, hybrid and dynamic stochastic ground-holding results for GDP Case 1 (Low duration of capacity deterioration).

	Static	Hybrid	Dynamic
Total ground-delay costs	210	211	185
Total airborne-delay costs	6.75	1.25	0

Table 6.2: Static, hybrid and dynamic stochastic ground-holding results for GDP Case 2 (Medium duration of capacity deterioration).

	Static	Hybrid	Dynamic
Total ground-delay costs	222.75	220.85	213.7
Total airborne-delay costs	0	0	0

Table 6.3: Static, hybrid and dynamic stochastic ground-holding results for GDP Case 3 (High duration of capacity deterioration).

As expected, the dynamic model solution achieves the lowest total delay costs across all three cases, followed by hybrid and static models in that order. The disparity in the optimal costs is larger for case 1 (shortest expected duration of deteriorated capacity) and smaller for case 3 (longest expected duration of deteriorated capacity), respectively. This trend in the pre-CDM gains of the dynamic model is consistent with the observations in Mukherjee and Hansen (2007) [55].

6.3.1.3.1 Post Intra-airline substitution The above-presented delay costs for slot allocation results assume cost homogeneity across all flights. Through parameter P2 listed above, variability is introduced in the unit delay costs across flights. The airlines would then employ intra-airline slot substitution to further reduce total delay costs beyond that of the slot allocation. Figures 6-3, 6-4 and 6-5 below present the total delay costs before and after intra-airline substitution for airline 1 for the three capacity cases. Each graph

provides separately the results for each stochastic model for different percentage coefficients of variation in unit delay costs (parameter P2), across the 25 flights operated by airline 1. Each cost value plotted in the graph is an average measure computed from a randomly generated sample set of 100 realizations of unit delay costs across the 25 flights, assuming that the unit delay costs are uniformly distributed about the mean with the specified variance. The same sample set of cost realizations is used for computing all three model results for a given variance measure.

Figure 6-3: Delay costs for airline 1 pre- and post- intra-airline substitution (Case 1 - Low expected duration of deteriorated capacity).

The results illustrate the previously mentioned tradeoff in delay cost reduction between the three stochastic models within the CDM framework. In all three capacity cases, the dynamic slot allocation has the lowest delay costs *before* the intra-airline slot substitution process. However, the superiority of the dynamic model *after* intra-airline slot substitution progressively decreases with increasing variability in flight delay costs, as the hybrid and static solutions experience higher gains through the slot substitution process. The crossover in terms of final delay costs (around 15-20 % variation for capacity cases 1 and 2) occurs at measures of cost heterogeneity that can be realistically expected for actual airline operations.

Figure 6-4: Delay costs for airline 1 pre- and post- intra-airline substitution (Case 2 - Medium expected duration of deteriorated capacity).

Building upon the trend in pre-CDM gains of the dynamic model discussed in the previous section, it is observed that capacity cases 1 and 3 are most and least favorable, respectively, for the dynamic model performance in terms of final delay costs.

Section 6.3.2 seeks to validate the inferences drawn from hypothetical case studies using real-world data for airline schedules and delay costs.

6.3.2 Real-world case studies

The comparative study of the stochastic ground-holding models is now extended to data from a GDP recorded on Feb 17, 2006 at LGA.

6.3.2.1 Base data

The GDP was in effect for 17 hrs from 7 am to midnight, during which a total of 542 arrivals, operated by 27 airlines, were scheduled. This tally excludes foreign arrivals, which are typically exempt from GDPs. The original arrival schedule prior to the issuance of the
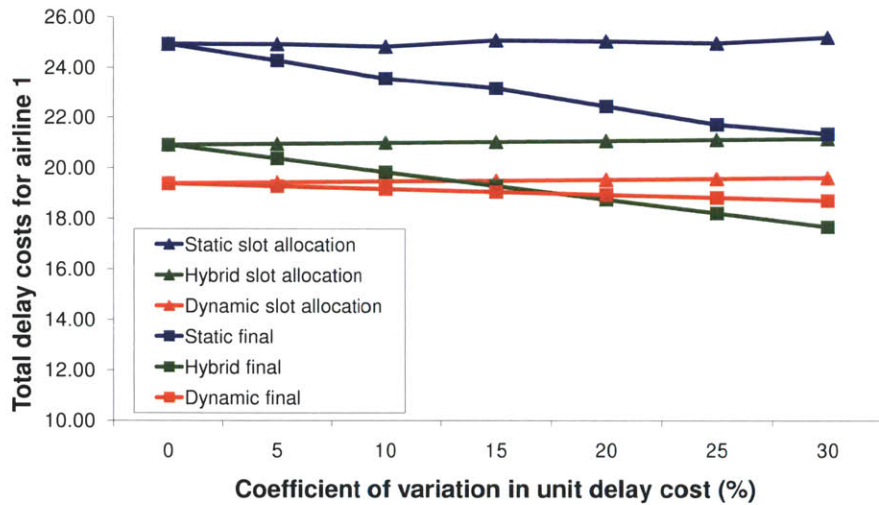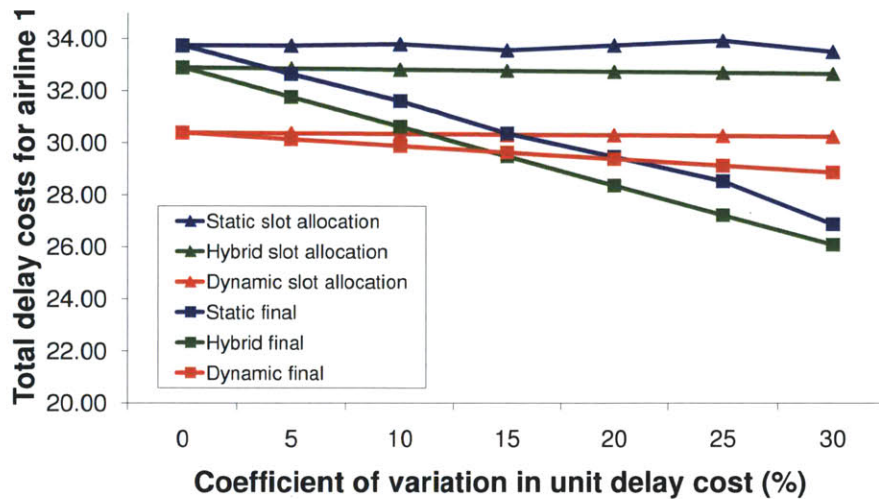
170
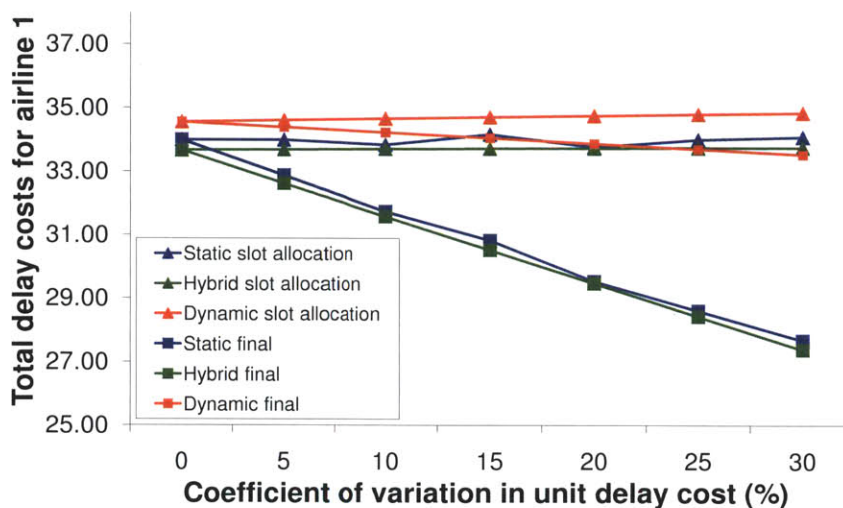
Figure 6-5: Delay costs for airline 1 pre- and post- intra-airline substitution (Case 3 - High expected duration of deteriorated capacity).

GDP was obtained from the ASPM 15-min dataset.

In this analysis, a discrete time interval is taken to be 1 hr long, and accordingly the duration of every flight is rounded up to the nearest hourly measure. The maximum flight duration among the domestic arrivals scheduled within the GDP time horizon was 5 hrs. Tables 6.4 and 6.5 give the hourly aggregate scheduled demand and a break-up of this demand by flight duration, for each airline.

Hour	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Demand	4	35	36	29	31	39	33	36	40	32	34	37	40	30	44	32	10

Table 6.4: Hourly aggregate scheduled demand (0700-2400 on Feb 17, 2006).

The 15-min arrival capacity estimated in Chapter 2 for LGA is scaled up to an hourly measure of 56 operations. Due to the lack of reliable data on capacity forecasts used during this GDP, synthetic scenario trees similar to that in the hypothetical case studies (Figure 6-2) are generated, with the capacity halved (28 arrivals/hr) under the deteriorated state.

Representative estimates for unit ground and airborne delay costs for each flight were obtained using a combination of publicly available sources on reported airline operating costs,

171

Airline (IATA code)	1-hr	2-hr	3-hr	4-hr	5-hr
AAL	0	3	37	11	12
DAL	28	8	24	2	0
EGF	3	51	2	3	0
CHQ	5	46	1	0	0
USA	31	12	3	0	0
CJC	10	29	0	0	0
COM	0	25	10	4	0
PDT	5	26	1	0	0
NWA	0	9	0	10	0
UAL	1	0	7	6	0
ACA	0	16	0	0	0
NKS	0	5	9	0	0
TRS	1	9	4	0	0
COA	0	4	0	0	9
AMT	0	0	9	0	0
AWI	1	6	0	0	0
MEP	0	0	4	3	0
JBU	0	0	6	0	0
ASH	1	4	0	0	0
JIA	1	4	0	0	0
LOF	0	5	0	0	0
BTA	0	4	0	0	0
EJA	1	0	1	2	0
CAA	0	3	0	0	0
JZA	0	3	0	0	0
CJA	0	1	0	0	0
EFG	0	1	0	0	0

Table 6.5: Number of flights per flight duration, by operating airline.

including T-100 schedules and P-52 files from the BTS website. The delay costs for a flight were assumed to comprise of three additive components: fuel, crew and passenger delay costs. The sources mentioned above were used to compute each of these cost components for each flight in the case study dataset. Details of the methodology used for flight-specific delay cost computation are provided in Appendix III.

Table 6.6 provides summary statistics for the computed unit delay costs by each operating airline, and for the entire dataset.

Airline	Unit ground delay cost ($/hr)		Unit airborne delay cost ($/hr)	
(IATA code)	Mean ($\times 10^3$)	Coeff. of Var.(%)	Mean ($\times 10^3$)	Coeff. of Var.(%)
AAL	4.1	16.7	8.82	11.4
DAL	4.17	24.2	9.66	14.2
EGF	1.45	6.8	3.78	2.6
CHQ	1.45	11.8	3.73	4.6
USA	2.87	13.9	7.19	5.5
CJC	2.24	56.1	5.24	59.8
COM	2.26	28.5	6.3	34.8
PDT	1.17	27.6	2.74	12.8
NWA	4.0	18.9	8.84	19.5
UAL	3.87	9.7	8.6	8.5
ACA	3.35	9.9	7.96	6.7
NKS	4.39	16.2	9.48	7.3
TRS	3.07	10.4	6.99	5.0
COA	3.43	18.6	7.87	9.2
AMT	4.34	2.3	9.74	5.2
AWI	1.31	12.7	3.82	4.3
MEP	3.06	0.8	8.02	2.2
JBU	4.61	0.8	8.56	0.5
ASH	2.93	35.9	7.00	34.4
JIA	1.25	10.5	2.23	6.1
LOF	1.47	5.7	3.96	2.1
BTA	1.36	8.6	2.82	3.8
EJA	3.82	24.2	9.36	7.1
CAA	1.6	0	3.7	0
JZA	3.0	0	8.7	0
CJA	3.2	0	7.3	0
EFG	1.3	0	3.16	0
Aggregate	2.81	47	6.6	41.2

Table 6.6: Summary statistics for unit delay costs for airlines at LGA.

6.3.2.2 Experimental design

The goal here is to validate the results from the hypothetical case studies using real-world data on demand and unit flight delay costs. To this effect, the influence of the following capacity-related parameters on the comparative performances of the three stochastic models on real-world data are studied.

P1 Expected duration of deteriorated capacity.

P3 Total length of GDP planning horizon (maximum of 17 hrs).

Here, parameter P3 is a newly introduced parameter, while P1 is as described for the hypothetical case studies in Section 6.3.1.

In order to generate a range of values for expected low capacity duration (P1) for a given GDP horizon length L (P3), a sequence of (2L-1) scenario probability distributions are generated in the following manner:

$$
\begin{aligned}
(L \text{ distributions}) \ \forall t \in \{1, .., L\} : p_s^t &= \begin{cases} \frac{1-0.01(L-t)}{t} & \forall s \in \{1, ..t\}, \\ 0.01, & \forall s \in \{t+1, .., L\} \end{cases} \\
(L-1 \text{ distributions}) \ \forall t \in \{1, .., L-1\} : p_s^t &= \begin{cases} 0.01 & \forall s \in \{1, ..t\}, \\ \frac{1-0.01t}{L-t} & \forall s \in \{t+1, .., L\} \end{cases}
\end{aligned}
\tag{6.6}
$$

The following sub-section presents the delay costs after intra-airline substitution for each stochastic model across this sequence of scenario tree cases, for GDP horizon lengths of 7, 10 and 15 hrs. For each GDP horizon length L, the relevant arrival schedule and unit delay costs data from 0700 hrs to (0700+L) hrs are extracted from the LGA dataset, and are applied to the sequence of capacity scenario trees described above.

6.3.2.3 Results

Figures 6-6, 6-7 and 6-8 summarize the final delay costs, aggregated over all airlines, for each stochastic model across the previously described range of values for test parameters (i.e., P1 and P3). The figures present the net % improvement achieved by each model over the worst stochastic model (as measured by the final aggregated delay costs).

For each GDP horizon length, the static model progressively takes over from the dynamic model as the model producing the lowest final delay costs as the expected duration of deteriorated capacity increases. Also, for the range where the dynamic model performs best, its percentage improvement over the worst model (typically static) is greater for higher GDP horizon lengths. Both these trends can be attributed to greater pre-CDM gains for the dynamic model under the specified ranges for the two parameters. As mentioned in the hypothetical case studies, the dynamic model's ability to produce superior pre-CDM gains for lower expected durations of deteriorated capacity was highlighted in Mukherjee and Hansen (2007) [55]. The pre-CDM gains for the dynamic model also grow with the length of GDP planning horizon, as a longer planning horizon implies more dynamic information on capacity scenario can be acquired and utilized by the dynamic model.
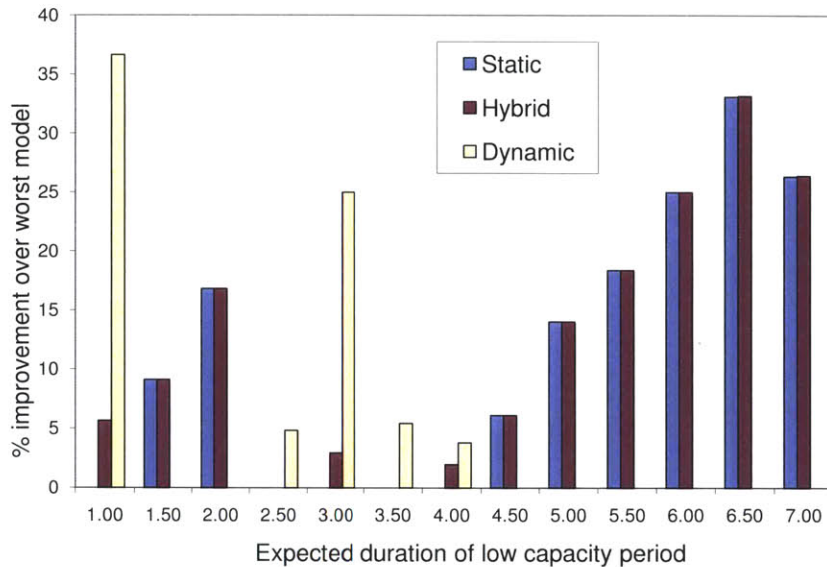
Figure 6-6: Percentage improvement in final system delay costs post intra-airline substitution (GDP horizon length = 7 hrs). The missing bar corresponds to the worst model.

It is also worth noting that the hybrid model is rarely the worst-performing model across the explored ranges of the two test parameters. This validates the underlying principle of the hybrid model, which is to combine the dynamic model's superior pre-CDM performance and the static model's greater amenability to intra-airline substitution.

Lastly, the influence of parameter P2 (variability in unit delay costs) on relative performance of the stochastic models is validated by focusing on individual airlines. From Table 6.6, airlines 'DAL' (coefficient of variation = 19%) and 'AAL' (coefficient of variation = 14%) are selected because of their traffic volumes. Figures 6-9 and 6-10 present the final, post- intra-airline substitution costs for these two airlines when GDP horizon length was set to 10 hrs.

The dynamic model produces the best results in case of AAL across all values of expected low capacity duration, while there is a cross-over point for DAL following which the static and the hybrid perform better. This is also in agreement with the results from the hypothetical case studies, where it was demonstrated that higher variability in unit delay costs (19% for DAL vs 14% for AAL) tended to favor the static and hybrid models over the dynamic model.

175

Figure 6-7: Percentage improvement in final system delay costs post intra-airline substitution (GDP horizon length = 10 hrs). The missing bar corresponds to the worst model.

The intra-airline slot substitution analysis shows that the hybrid model is seldom the worst-performing among the three stochastic models, across the range of GDP cases studied using hypothetical and real-world data. This conclusion supports the objective behind the hybrid model, which was to effectively reconcile the tradeoffs between the static and dynamic stochastic models in a CDM environment. The next section extends this comparative analysis to the SCS/Compression mechanism.

6.4   Slot Credit Substitution (SCS)

SCS is a real-time, adaptive form of Compression, and is currently viewed as its long-term replacement. The traditional compression mechanism is a batch process executed at periodic intervals during a GDP, simultaneously handling multiple slot forfeitures from airlines by advancing flights to occupy vacated slots. SCS is an asynchronous, event-driven version of compression triggered by a single forfeiture request from an airline. The dynamic responsiveness of the SCS helps it further improve GDP efficiency compared to compression.
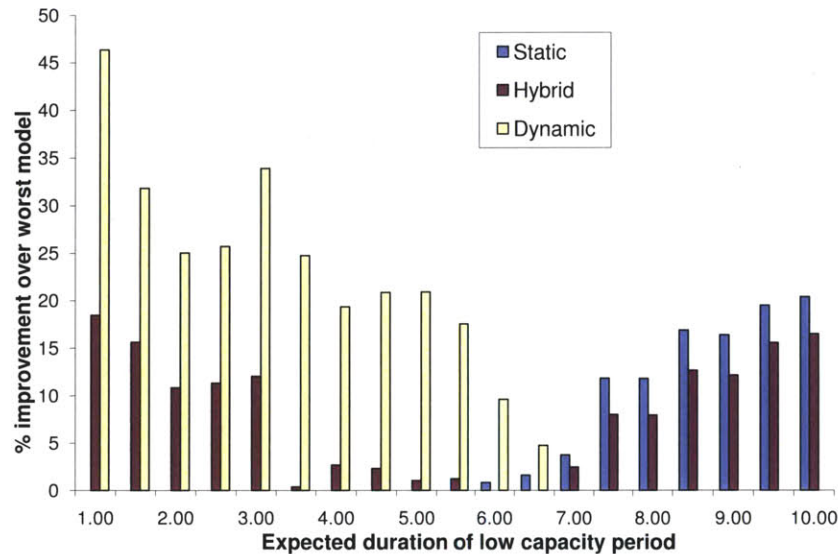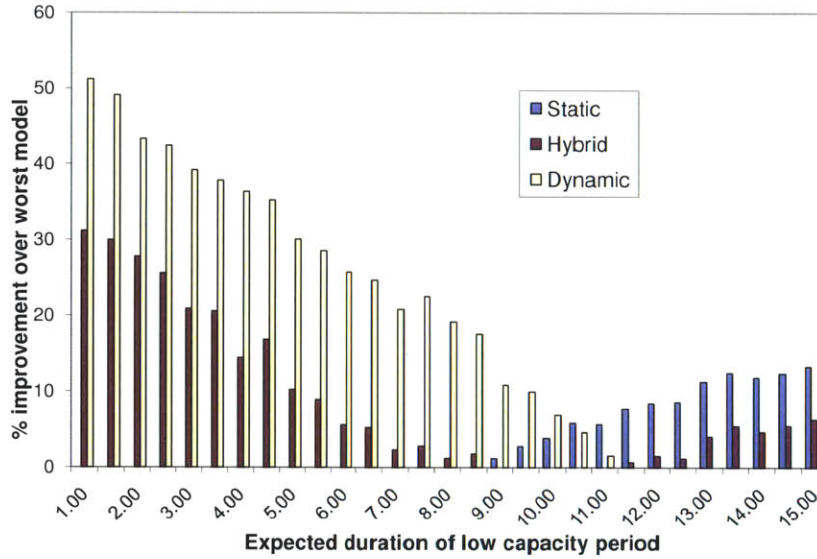
Figure 6-8: Percentage improvement in final system delay costs post intra-airline substitution (GDP horizon length = 15 hrs). The missing bar corresponds to the worst model.

A detailed account on the SCS mechanism is provided in Howard (2002) [41], and its potential advantages are highlighted in Ball et. al. (2005) [3]. In our study, the SCS mechanism is considered instead of Compression due to its ever-growing prominence.

Chapter 1 presented an illustrative example of SCS/Compression in its current deterministic design. The reader is directed to Chang et. al. (2001) [18] for more details regarding the same. Listed below are three features that summarize the functioning of a SCS within the context of a stochastic slot allocation.

1. **Slot forfeiture compensation:** An airline wishes to forfeit slot $slot_c$ (originally assigned to the delayed/cancelled flight c) while seeking a compensatory slot with a later arrival time k across all scenarios. This later arrival time might either correspond to the earliest arrival time for flight c following mechanical delays, or the earliest arrival time for another flight operated by the given airline.

2. **Pareto delay improvement to other flights:** When executing SCS or Compression on a deterministic slot allocation, all flights except the delayed/cancelled flight c are

177
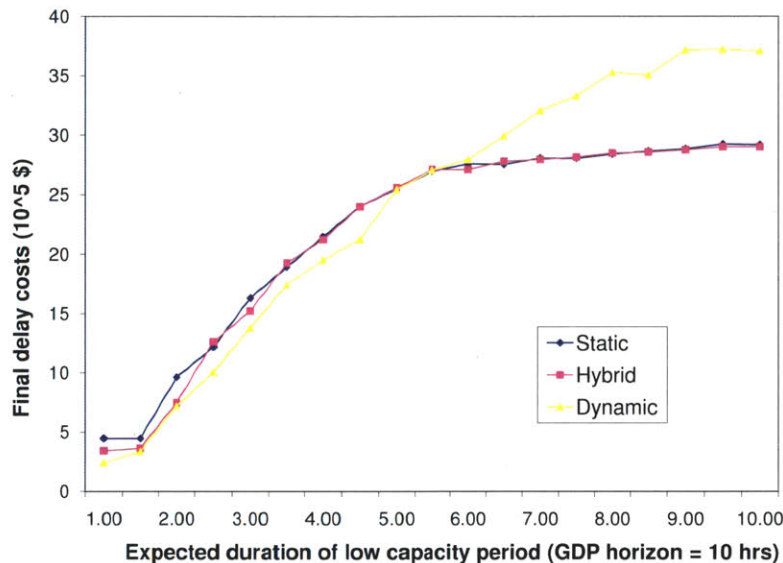
Figure 6-9: Final delay costs for DAL (GDP horizon length = 10 hrs).

either advanced or retained at their assigned slot timings. With respect to a stochastic slot allocation, an equivalent feature would be to ensure these other flights are either advanced or retained at their current slot timings *for every scenario*. This "pareto delay improvement" feature was also adopted by Mukherjee and Hansen (2007) [55] in their formulation of the compression mechanism, which was applied to dynamic stochastic slot allocation.

3. **Equitable distribution of delay improvements** Apart from pareto improvements to other flights, SCS or Compression also ensures that these improvements are distributed uniformly among all candidate flights. For a simple illustration of this feature in a deterministic setting, consider the following 3-flight example with their respective slot allotments and earliest times of arrival as given in Table 6.7.

Now suppose flight A experiences a mechanical delay of 1 hr (changing its ETA to 0800 hrs), and the airline operating it is willing to forfeit its current slot (at 0720 hrs) in exchange for one at 0800 hrs. There are two possible rearrangements to the current slot allocation that would achieve this, as given below in tables 6.8 and 6.9.

178
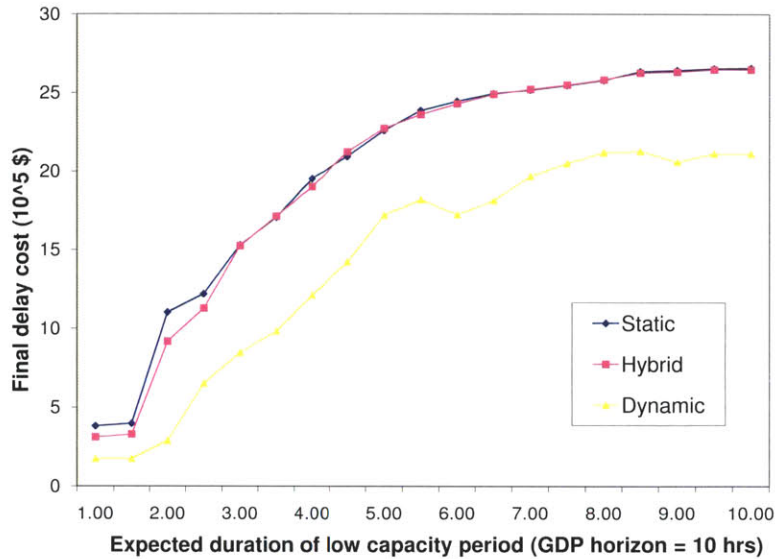
Figure 6-10: Final delay costs for AAL (GDP horizon length = 10 hrs).

Flight ID	ETA	Allotted slot timing
A	0700	0720
B	0710	0740
C	0720	0800

Table 6.7: GDP slot allocation for 3-flight schedule.

Between these two candidate solutions, the principles of compression or SCS dictate that solution 2, involving slot advancements to flights B and C, be chosen over solution 1, involving slot advancement to flight C alone. This feature, when translated to the context of a stochastic slot allocation, would imply that the number of flights advanced for every scenario when accommodating a slot forfeiture needs to be maximized.

A formulation for executing SCS on stochastic slot allocations is developed that incorporates all of the above three features. The input for this SCS formulation comprises of two components:

1. Scenario-specific stochastic slot allocation solution (in richer detail than that communi-

Flight ID	ETA	Allotted slot timing
C	0720	0720
B	0710	0740
A	0800	0800

Table 6.8: Candidate SCS solution 1.

Flight ID	ETA	Allotted slot timing
B	0710	0720
C	0720	0740
A	0800	0800

Table 6.9: Candidate SCS solution 2.

cated to airlines as described in section 6.2.2).

$\text{arr\_t}^s(\text{slot}_f)$: allotted arrival time under scenario s for slot_f assigned to flight f

$Aq_{s,t}^{\text{orig}}$: arrival queue lengths under scenario s for time t

$\text{dur}(\text{slot}_f)$: duration of flight assigned to slot_f

$\text{ETA}(\text{slot}_f)$: earliest time of arrival for flight assigned to slot_f.

As mentioned earlier, the first two attributes do not change through the intra-airline slot substitution process, assuming no cancellations have occurred apart from the SCS triggering request.

2. SCS triggering request: Airline forfeiting slot slot_c, currently occupied by flight c, in exchange for a later slot at or beyond time k.

The formulation for the SCS mechanism given above inputs is presented below.

180

$$\text{Maximize} \quad \sum_{s \in S} p_s [(\sum_{f \in F \setminus c} \sum_{r=1}^{T} \text{savng}_r Y_{f,r}^s) - M d_c^s] \tag{6.7}$$

subject to:

$$\sum_{t=\text{ETA(slot}_f)}^{\text{ETA(slot}_f)+K} X_{f,t}^s = 1, \ \forall s \in S, \forall f \in F \tag{6.8}$$

$$Aq_{s,t} \geq \sum_{f \in F} X_{f,t}^s + Aq_{s,t-1} - A_{s,t}^{\text{cap}}, \ \forall t \in \{1,..,T\}, s \in S \tag{6.9}$$

$$X_{f,t}^{s^1} = X_{f,t}^{s^2}, \ \forall s^1, s^2 \in G(t - \text{dur(slot}_f)), \tag{6.10}$$

$$d_c^s = \sum_{t=1}^{T} t X_{c,t}^s - k, \ \forall s \in S, \tag{6.11}$$

$$\sum_{t=1}^{T} t X_{f,t}^s \leq \text{arr\_t}^s(\text{slot}_f), \ \forall s \in S, f \in F \setminus c \tag{6.12}$$

$$Aq_{s,t} \leq Aq_{s,t}^{\text{orig}}, \ \forall s \in S, t \in \{1,..,T\} \tag{6.13}$$

$$Y_{f,r}^s = X_{f,\text{arr\_t}^s(\text{slot}_f)-r+1}^s, \ \forall s \in S, f \in F \setminus c, r \in \{1,..,T\} \tag{6.14}$$

$$X_{f,t}^s \in \{0,1\}, \ d_c^s \geq 0, \ \forall t \in \{1,..,T\}, \forall s \in S, \forall f \in F$$

Notations (in addition to that previously described):

Decision Variables

$Y_{f,r}^s$: 1 if flight assigned to slot sl_f is advanced by r time units under scenario

s, 0 otherwise

d_c^s : Ground delay beyond k for delayed/cancelled flight c under scenario s

Objective function coefficients

savng_r : Measure of benefit for advancing a flight by r time units

M : A very large number (around 10000).

Constraints

(6.8) : Arrival demand balance

(6.9) : Arrival queue balance

(6.10) : Non-anticipativity on arrival ground-holding

181

(6.11) : Constraint ensuring flight c is allotted a compensatory slot at or beyond time step k

(6.12) : Constraint ensuring every flight apart from c is either advanced or retained in its current arrival time across all scenarios

(6.13) : Constraint ensuring arrival queue length is not increased for any time interval across any scenario

(6.14) : Constraint measuring the advancement in time units for any flight apart from c

Note that constraint (6.11), in conjunction with the second component of the objective function (6.7) (involving the parameter M), enforces the first feature of SCS pertaining to slot forfeiture compensation. The airline operating flight c is allotted a compensatory slot as close to, but not earlier than the specified time interval k across all scenarios.

Next, the combination of the two constraints (6.12) and (6.13) serve as proxies for the second SCS feature pertaining to pareto improvement of other flights. This approach is similar to the one adopted in Mukherjee and Hansen (2007) [55], and designed to ensure that no flight is worse off in terms of ground delay (via constraint (6.12)) or airborne delay (via constraint (6.13)) for any scenario compared to its original allocation.

Finally, the first component of the objective function (6.7) featuring the variable $Y_{f,r}^t$, as defined by constraint (6.14), coupled with the "benefits" function $savng_r$ being increasing but strictly concave, will enforce the third feature of SCS pertaining to equitable distribution of delay benefits. To understand how this works, reconsider the 3-flight illustrative example given above. Assuming the allotted slots occupy time intervals 1,2 and 3, the first component of the objective function (6.7) would measure $savng_2$ for candidate SCS solution 1 (6.8), and $2 \times savng_1$ for candidate SCS solution 2 (6.9) respectively. Given $savng_r$ is concave, we know that $2 \times savng_1 > savng_2$, which renders candidate solution 2 preferable.

The above formulation has been customized for dynamic stochastic model through the non-anticipativity constraint (6.10). The equivalent constraints when SCS is executed on

the static and hybrid stochastic solution, respectively, are:

$$X_{f,t}^{s^1} = X_{f,t}^{s^2}, \ \forall s^1, s^2 \in S \qquad \text{for Static model} \qquad (6.15)$$

$$X_{f,t}^{s^1} = X_{f,t}^{s^2}, \ \forall s^1, s^2 \in G(\text{ETA}(\text{slot}_f) - \text{max\_dur}) \qquad \text{for Hybrid model} \qquad (6.16)$$

This customization ensures the key properties of the slot allocation from a given stochastic ground-holding model are preserved following the execution of SCS. Thereby, the format of intra-airline slot substitutions, which depends upon the properties of the slot allocation as discussed in the previous section, is not altered post-SCS.

Section 6.4.1 discusses the application of the above-described SCS formulations to the slot allocations from the three stochastic models for the hypothetical cases, and compares the results across the three models using relevant metrics. Since SCS does not require flight-specific information beyond that available to the airport (i.e, dur(slot$_f$), ETA(slot$_f$), etc.), real-world data is not required to validate the findings from the hypothetical case studies.

6.4.1 Hypothetical case studies

6.4.1.1 Base Data

The same base data from the hypothetical analysis of intra-airline slot substitutions, namely, the capacity scenario tree, arrival demand and unit delay cost coefficients, are adopted here.

6.4.1.2 Experimental Design

The GDP planning horizon is set to 10 intervals. A sequence of scenario tree probability distributions as described in expression (6.6) is generated. For each scenario tree case, a random sample of 100 SCS requests is generated out of the 138 scheduled flights. Each SCS request specifies a pair of forfeited slot slot$_c$ and compensatory slot at time k. These SCS requests are processed for the slot allocation solution from each stochastic model using the corresponding SCS formulation ((6.15) for static, (6.16) for hybrid and (6.10) for dynamic).

The objective here is to compare the benefits from the SCS mechanism between the three stochastic models. To this end, two classes of performance metrics are defined.

1. **Delay costs for delayed/cancelled flight c:** The expected ground (G_c) and airborne (A_c) delay costs experienced by flight c across all scenarios following its reallocation based on its new earliest time of arrival k. Lower values for G_c and A_c are preferable, and thus equate to higher benefits from the SCS mechanism.

2. **Delay benefits for other flights (B_0):** The total expected reduction in delay (ground + airborne) costs across all scenarios for all other flights. High value for B_0 indicates higher benefits from the SCS mechanism.

These two metrics together add up to the system delay benefits (B_s) obtained from SCS, as described below:

$$B_s = B_0 - G_c - A_c. \tag{6.17}$$

Note that all three metrics G_c, A_c and B_0 are computed after obtaining the post-SCS slot allocation. Given that the flexibility for swapping slots between successive flights varies between the static, hybrid and dynamic solutions, not all intervening flights between the forfeited and compensatory slots need be feasibly advanced across all scenarios for a given SCS request. This means that the same SCS request might generate different values for the metric B_0 (delay benefits for other flights) for the three stochastic models. This feature also implies that a compensatory slot need not be available at the specified later time slot k across all scenarios, resulting in different values for the metric G_c for the three stochastic models. It is the nature of these potential discrepancies in SCS benefits between the static, hybrid and dynamic models that is of interest in the following results.

6.4.1.3 Results

The figures presented below plot the average values for the above-defined metrics computed over the respective sample of SCS requests for each scenario tree case. Figures 6-11, 6-12, 6-13 and 6-14 depict the results for metrics G_c, A_c, B_0 and B_s, respectively. For interpreting the delay cost values presented, recall that the unit ground and airborne delay costs used in the stochastic ground-holding model formulations were 0.5 and 2.5 respectively.

Discussed below are the notable trends pertaining to each performance metric inferred from the results.
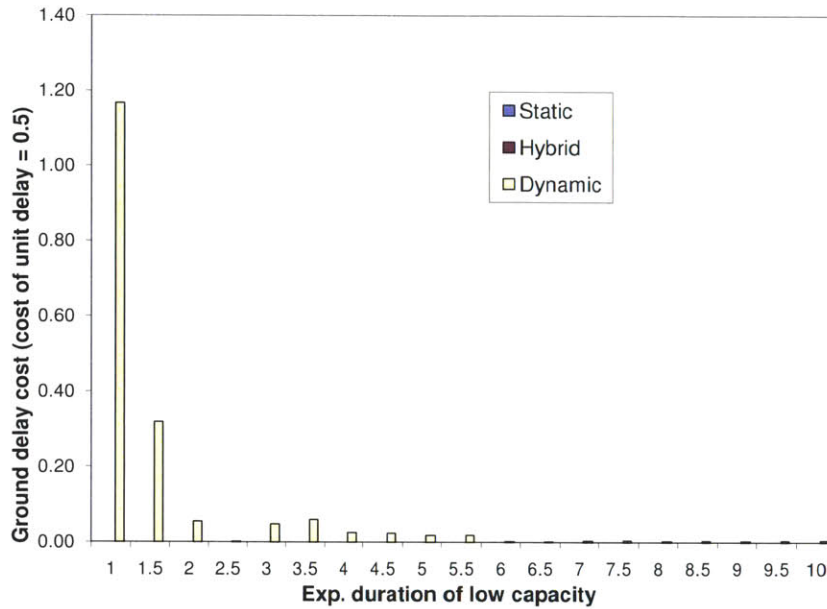
Figure 6-11: Average ground delay costs for cancelled/postponed flight, G_c (GDP horizon length $= 10$).

G_c: Across all GDP cases, no SCS request results in any incremental ground delay for the delayed/cancelled flight when applied to the static or hybrid model (figure 6-11). This implies that, for these two models, a compensatory slot is always found at the specified later time by advancing other flights in the slot allocation solution. The same is not true for the dynamic model, especially at low values for expected duration of diminished capacity. Therefore, for the slot allocation from the dynamic model, a compensatory slot cannot always be provided at the specified later time of a SCS request. This is due to the reduced flexibility for swapping slots between successive flights in the dynamic model's slot allocation, as previously mentioned.

A_c: Another consistent trend across all GDP cases is that the incremental airborne delay for the delayed/cancelled flight is greater for static and hybrid models (figure 6-12). This is due to the fact that the static and hybrid models' slot allocations inherently suffer greater airborne delays compared to that of the dynamic model (refer tables 6.1, 6.2 and 6.3). This translates to greater airborne delays for individual slots. Therefore,

185
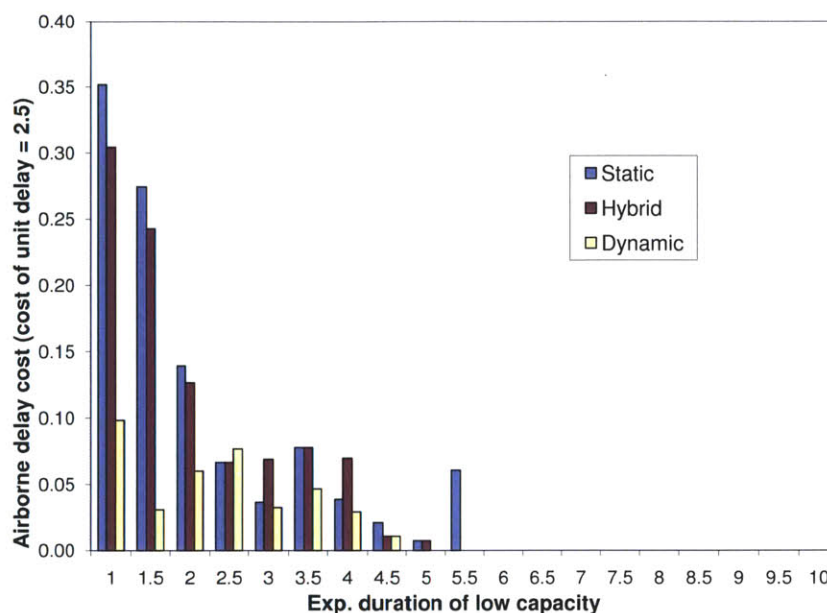
Figure 6-12: Average airborne delay costs for cancelled/postponed flight, A_c (GDP horizon length = 10).

the compensatory slot for the SCS request, while alloted at the specified later time without further ground delay, can be subject to greater airborne delay under the static and hybrid's slot allocations.

B_0: In terms of the delay benefits generated for the other flights, the observed trend across the GDP cases is largely driven by two counteracting effects. Firstly, the dynamic model, by virtue of its superior pre-CDM allocation, can potentially assign earlier arrival times for most flights in the original slot allocation, including the delayed/cancelled flight c. This would typically imply a greater number of intervening slots between the forfeited slot and the specified later time, enabling greater aggregate delay benefits from advancing the flights in these intervening slots, as observed for some of the GDP cases in Figure 6-13.

However, these benefits cannot always be realized for the dynamic slot allocation due to the above-mentioned limitation in the flexibility for advancing these intervening slots. As described above, this limitation also results in incremental ground delay D_c

186
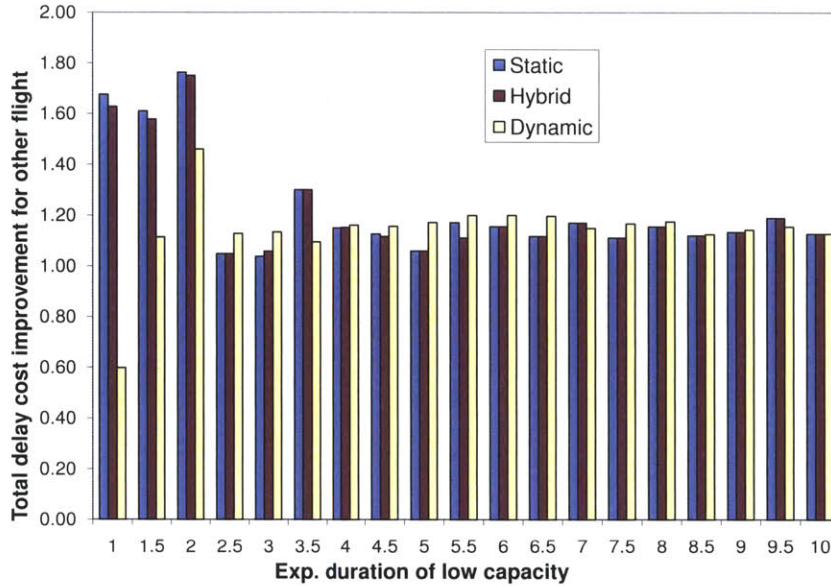
Figure 6-13: Average total delay benefits for other flights, B_0 (GDP horizon length $= 10$).

to the delayed/cancelled flight upon reallocation. As observed in the plotted figures, the GDP cases where the dynamic model realizes markedly lower delay benefits for the other flights (B_0 in Figure 6-13) are the same as those where the dynamic model realizes the most significant incremental ground delays for the delayed/cancelled flight (G_c in Figure 6-11).

B_s: The total system delay benefits B_s, as observed in figure 6-14, follows a similar trend to the metric B_0, and is a compact illustration of the underlying tradeoff between the static, hybrid and dynamic models. The dynamic model performs comparably or better than the static and hybrid models at higher values for expected duration of diminished capacity, while it seems to perform significantly worse for GDP cases with low expected duration of diminished capacity. The performances of the static and hybrid models are almost identical across the entire range of GDP cases.

In summary, it is noted that, across a range of GDP input cases, there is tangible evidence of tradeoff between the static, hybrid and dynamic models' performances in terms of the realized benefits from the SCS mechanism. Neither model consistently outperforms the

187
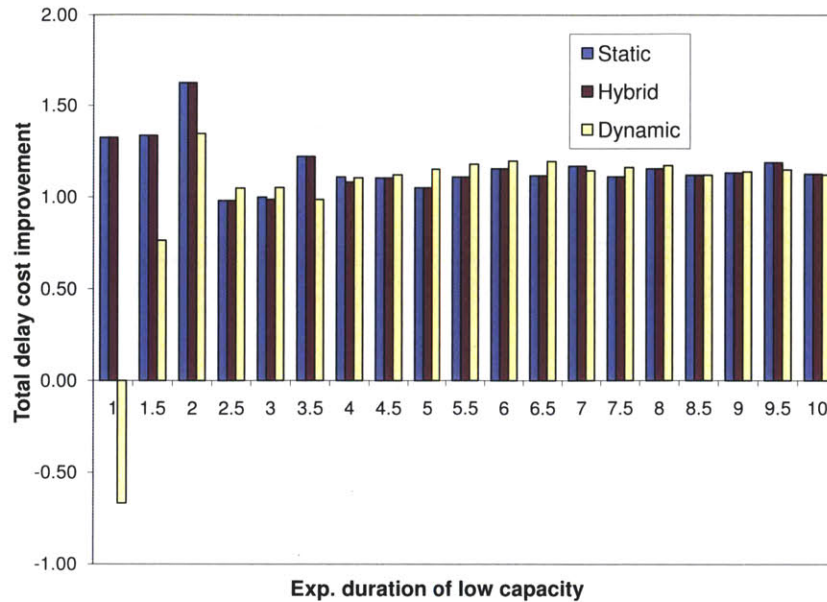
Figure 6-14: Average total delay benefits for all flights, B_s (GDP horizon length = 10).

others across the range of GDP cases studied. This inference corresponds with the one derived from the intra-airline substitution analysis regarding the ability of the hybrid model to be better than the worst-performing model across the studied range of GDP inputs.

This chapter focused on three single-airport stochastic ground-holding models discussed in the previous chapter: Static, Hybrid and Dynamic. These three models were examined within the context of contemporary ground delay programs (GDP) which involve collaborative decision making (CDM). The CDM paradigm implements two mechanisms that allow airlines to participate in the GDP: intra-airline substitution and compression. Recently, the mechanism of Slot Credit Substitution (SCS) has been adopted as a more efficient version of Compression. The hybrid stochastic model was developed as an attempt to blend the pre-CDM efficiency of the dynamic model and the CDM amenability of the static model. The objective of this chapter was to examine the effectiveness of this blending achieved by the hybrid model through case studies combining the stochastic ground-holding models with the CDM mechanisms.

A conceptual extension to the GDP framework to support slot allocation solutions from

stochastic ground-holding models was designed first. Stochastic equivalents for intra-airline substitution and SCS mechanisms as applicable to the stochastic slot allocations were formulated. The performances of the three stochastic models with respect to the two CDM mechanisms were compared through GDP cases constructed using both hypothetical and real-world data. Appropriate performance metrics were defined with respect to each CDM mechanism to facilitate this comparison. With respect to intra-airline substitution, it was demonstrated how the hybrid model effectively bridges the disparity in performances between the static and dynamic models over a range of GDP input cases. Over a similar range of GDP cases, the tradeoff between the three models in terms of system delay benefits from the SCS mechanism were also illustrated.

Chapter 7

Conclusions

7.1 Thesis Summary

This thesis addressed the problem of efficiently allocating airport arrival and departure capacity in the presence of uncertainty. In particular, it addressed limitations in the current design of the Ground Delay Program (GDP), a prominent congestion management scheme at airports facing short-term capacity reduction. Present-day GDPs are executed within a collaborative decision-making (CDM) paradigm in which airlines use flight-specific costs to revise the delay allocation. However, the airport capacity is assumed to be deterministic and is allocated only among arrivals within the GDP framework. This necessitates the revised execution of GDPs with every updated capacity forecast, while departures are handled in an ad hoc manner. Historical data on GDPs suggest there exists significant potential for system-wide delay benefits that can be realized by correcting these limitations. While prior studies in literature examine the two issues separately, this thesis considered them in tandem by proposing refinements of the GDP algorithms.

First, a comprehensive modeling framework was developed for characterizing key airport operational elements and their influence on arrival and departure capacities, which are interdependent quantities. An approach based on quantile regression was designed for empirically estimating airport capacity envelopes using observed throughputs. This novel approach explicitly quantified impacts of exogenous factors like configuration, visibility, etc. on airport capacity, and used two effective criteria to identify and eliminate outliers from the observa-

tion dataset. The proposed methodology was applied to the NY airspace system comprising of LGA, EWR and JFK airports, and was used to estimate intra- and inter-airport capacity envelopes and their dependence on influencing factors. The latter set of envelopes captured the metroplex effects, that is, the inter-dependence of operational capacities across neighboring airports that share terminal airspace. The thesis also studied the dynamics of runway configuration, which is a key determinant of airport capacity. The selection of active runway configuration at a given time step was modeled using a discrete choice framework, and applied to EWR and LGA to extract the underlying drivers for this decision process, using recorded airport observations. The same observations were also used to uncover the impacts of configuration switches on airport efficiency through a two-stage regression approach.

Following the airport capacity estimation module, the thesis addressed capacity allocation in GDPs by simultaneously accounting for forecast uncertainty, and the joint handling of arrivals and departures. Arrivals-only stochastic ground-holding models from literature were extended to include departure considerations. Integrated versions of two standard stochastic models, *static* and *dynamic*, were applied to case studies featuring hypothetical and real-world GDP data to highlight the advantages over respective arrivals-based approaches. The results indicated that typical GDP instances can benefit considerably in terms of system delay costs and equity by adopting an integrated approach to capacity allocation under uncertainty.

Finally, the application of CDM mechanisms to stochastic ground-holding solutions was analyzed. The relative amenability of the static and dynamic solutions to slot substitution, and the resulting tradeoffs between pre-CDM and CDM-induced delay benefits of these two models was established. A new hybrid stochastic ground-holding model was developed to combine the favorable features of the static and dynamic models, and its ability to achieve an effective compromise between the two was demonstrated through GDP case studies. To conduct the case studies, integer formulations were designed that extended the principles of Intra-airline slot substitution and Compression to the three stochastic ground-holding models. The results of the case studies highlighted the robustness of the hybrid stochastic model in terms of post-CDM delay metrics across a range of GDP inputs. In addition, the properties of the integer formulation for the hybrid stochastic ground-holding model was

studied, and two interesting results concerning its computational tractability were derived.

The following section highlights some of the compelling research directions that can help build upon the work presented in this thesis.

7.2   Future Research Directions

- **Incorporation of fleet mix and demand into capacity estimation:** Fleet mix and operational demand are two elements that were not explicitly considered in capacity estimation methodology proposed in Chapter 2. Given that the mandated separations between successive aircraft operations, be it arrivals or departures, are a direct function of the aircraft types, the composition of the aircraft fleet being served by an airport at a given time determines its throughput capacity. The accuracy of the parameter estimates obtained in Chapter 2 can be improved by designing an enhanced model specification that accounts for fleet size mix effects.

  In addition, empirical approaches for capacity estimation, including Gilbo (1993) [31], deal with an unfiltered dataset of airport observations that include instances where the operational demand was much less than capacity. The bias induced by such low throughput points can only be eliminated by identifying the set of observations when airport capacity was exceeded by demand. Recent efforts have been successful in extending the quantile regression approach to address this shortcoming through modeling advancements [70].

- **Uncertainty in capacity estimation:** The methods discussed in this thesis focus on deriving a deterministic measure of airport capacity. However, the realized airport capacity is the output of an interplay between various operational elements that inherently contain uncertainties, such as controller decisions, aircraft arrival and departure processes, etc. Some of these uncertainties are addressed by the models for configuration dynamics discussed in the Chapter 3. However, methodological refinements that directly model uncertainties in airport parameters within the capacity estimation framework can extract information from empirical data in a more reliable fashion.

- **Integrated modeling framework for configuration selection and switch impacts:** The models in Chapter 3 characterized runway configuration selection and configuration switch impacts in two independent steps. An interesting extension to this body of work would be to pursue an integrated statistical framework that estimates both operational elements. A conceivable option is a two-level decision framework, with the first level modeling configuration choice and second capturing the duration and operational impact of a configuration switch if determined at the first level. A similar modeling framework has been successfully adopted for the case of driver behavior modeling by Ahmed et. al. (1996) [2]. Such an integrated framework would extract richer information from available data during estimation since it considers the relation between configuration choice and switch effects.

- **Passenger and aircraft connectivity within integrated ground-holding:** The integrated stochastic ground-holding models presented in Chapter 4 did not consider passenger or aircraft connections between arrivals and departures. Since such information is not typically available to the system operator in advance, they cannot be enforced in the form of explicit constraints during ground-hold allocation. Given that the operating carriers may have alternative means of reallocating passengers or aircraft, the CDM framework offers the ideal setting to exercise such airline-specific decisions. Existing designs for CDM mechanisms of Intra-airline substitution and Compression only consider arrival slots. Future research should therefore explore integrated CDM mechanisms that would allow airlines to manage their arrival and departure slots simultaneously.

Some preliminary analysis has already been conducted on the topic of accommodating scheduled connections during a GDP. Gilbo (2000) [30] proposed acquiring information from airlines regarding priority flights amongst arrivals and departures for each time period that would be granted preference during ground-hold allocation. This provides airlines means of communicating their connectivity requirements. Hall (1999) [35] proposed allocating airport capacity as an integrated bundle of arrival and departure slots that would allow airlines to swap between arrival and departure slots without violating

capacity envelope limits. Airlines can thereby delay connecting departures in exchange for advancement in arrivals. The first method does not study potential incentives for airline to be truthful, and the second method is applicable to deterministic settings alone. Extensions to CDM mechanisms along the lines of that proposed in Hall (2009) [35] represent a promising future direction for stochastic models.

- **Revisions due to forecast updates for stochastic GDPs:** The stochastic GDP framework applied in the case studies of Chapters 4 and 6 inherently assumed that the parameters of the scenario tree forecasts, in terms of capacity values and scenario probabilities, were accurate. This might not always hold true, and subsequent forecast updates might require revisions to slot allocation and CDM mechanisms, as currently performed under the deterministic framework. This makes a case for comparing the pros and cons of the two frameworks through realistic GDP simulations involving forecast-driven revisions. While the deterministic framework permits scenario-specific flexibility in slot swapping unlike the stochastic framework, the number of revisions are likely to be fewer in the case of the latter, which ensures reliable information for airlines to make slot substitution/cancellation decisions. The experimental studies will essentially help answer the question of whether the airlines better off with a stochastic framework that will cause fewer and less drastic revisions to slot allocations, or a deterministic framework that will give rise to more significant revisions but allow them to swap slots specific to an updated scenario.

Appendix A

Hypothesis Testing in QR-based Estimation

Hypothesis tests are conducted to assess the statistical significance of empirically estimated values of model parameters, and is an important process for developing specifications of statistical models. Usually, every new explanatory factor appended to a model specification is subjected to a Nested Hypothesis test following its estimation. The general Nested Hypothesis test, in principle, attempts to evaluate the improvement in the statistical fit achieved by a model specification (M) over a restricted (nested) version (RM) of itself obtained by imposing a set of linear equality constraints $\mathfrak{L}$. $|\mathfrak{L}|$ is regarded as the number of degrees of freedom arrested in (M) to derive (RM). For the model with the new explanatory factor (M), the restricted model (RM) is obtained by fixing the coefficient of new factor to zero.

For quantile regression models, Koenker (2001) [46] develops the Nested Hypothesis tests based on likelihood ratios. The unrestricted model (M) is deemed statistically superior to the restricted model (RM) if the difference in their estimated likelihood ratios $L_M - L_{RM}$ exceeds a threshold value $T(\tau, |\mathfrak{L}|)$ which is a function of the estimated order of quantile τ and the degrees of freedom separating (RM) from (M), namely, $|\mathfrak{L}|$. For the piecewise-linear quantile function $Q_\tau^{dep}(y|x)$ defined for capacity envelopes in (2.2), the likelihood ratio test can be applied to assess the statistical significance of a new factor θ^{new} in the model if one can define the appropriate set of linear equalities $\mathfrak{L}$ to derive the restricted model (RM) that has coefficients for θ^{new} set to zero. The expression for the quantile function $Q_\tau^{dep}(y|x)$ from

197

(2.2) is reproduced here to assist subsequent discussion.

$$Q_\tau^{\text{dep}}(y|x) = \sum_i \alpha_k^i \theta^i + \left(\sum_i \beta_k^i \theta^i\right)x, \text{ for } (k-1) \leq x \leq k, \ \forall k \qquad (\text{A.1})$$

where y and x represent the departure and arrival counts respectively.

Given the piecewise-linear form for $Q_\tau^{\text{dep}}(y|x)$, each factor θ^i in the model specification has an intercept and a slope coefficient (α_k^i and β_k^i) for each linear piece $k \in \{1, ...x_{\text{max}}\}$, making it a total of $x_{\text{max}} * 2$ coefficients per factor. Returning to the earlier discussion, a straightforward way to set the coefficients of the new factor θ^{new} to zero (and thereby obtain the restricted model (RM) to perform the Nested Hypothesis test) is by defining a total of $x_{\text{max}} * 2$ linear equalities, explicitly restricting every coefficient. However, as the following example illustrates, this might be an overestimation of the actual degrees of freedom separating the restricted model (RM) from the unrestricted model (M) for typical estimates for $Q_\tau^{\text{dep}}(y|x)$.

Assume a model specification for $Q_\tau^{\text{dep}}(y|x)$ defined across 5 linear pieces ($x_{\text{max}}=5$), with one existing and one new explanatory factor θ^1 and θ^{new} respectively and the following estimated values for their coefficients. Now, there are a total of 10 coefficients estimated

k (Linear piece)	α_k^1	β_k^1	α_k^{new}	β_k^{new}
1	14	0	1	0
2	15	-1	1	0
3	15	-1	2	-0.5
4	15	-1	2	-0.5
5	20	-2	1	-0.5

Table A.1: Coefficients estimates for piecewise linear quantile function $Q_\tau^{\text{dep}}(y|x)$

over 5 linear pieces for the factor θ^{new} in this example. However, there is equality between coefficients of adjacent pieces in some cases (e.g. for $k=1$ and 2; $k=3$ and 4), which suggests that not all 10 coefficients represent independent degrees of freedom.

Consider the following set of linear equalities that can be imposed upon the above specification for quantile function $Q_\tau^{\text{dep}}(y|x)$ to restrict all 10 coefficients of factor θ^{new} to take

value 0.

$$\alpha_1^{\text{new}} = 0$$

$$\beta_1^{\text{new}} = 0$$

$$\beta_{i+1}^{\text{new}} = \beta_i^{\text{new}} \quad \forall i \in \{1, 2, 3, 4\} \tag{A.2}$$

It can be shown that adding these constraints to the linear programming formulation described in equations (B.1) - (B.5) ensures only factor (θ^1) has non-zero coefficients in the estimated result. Now, consider relaxing the following subset of linear equalities from the above-defined set (A.2).

$$\alpha_1^{\text{new}} = 0$$

$$\beta_3^{\text{new}} = \beta_2^{\text{new}} \tag{A.3}$$

By imposing those linear equalities from the set (A.2) not in set (A.3) on the linear programming formulation (B.1) - (B.5), one can recover the coefficient estimates given in Table A.1, i.e., the estimation results for the unrestricted quantile function specification.

This shows that the four linear equalities in set (A.3) are the only constraints that need to be imposed upon the unrestricted, 2-factor piecewise-linear model specification to obtain the restricted, 1-factor specification. Hence, the actual number of degrees of freedom separating the unrestricted and restricted models in the Nested Hypothesis test for the factor θ^{new} in the above example is 2 instead of 10.

This principle is adopted to compute the actual degrees of freedom when performing the Nested Hypothesis test (using likelihood ratios) for each new factor estimated for the airport capacity functions in Chapter 2.

Appendix B

Hypothesis Testing for Convexity of Capacity Envelope

Convexity of an airport capacity envelope has been implicitly assumed in most discussions in literature [31, 58, 71]. In this section, we demonstrate how the estimation framework described in Chapter 2 can be used to statistically verify this assumption.

A weaker but less arguable version of the convexity assumption is that of monotonic non-increment for the capacity envelope. This means that, in Figure 2-4, the departure capacity is a monotonically non-increasing function of arrival counts, i.e. the slope of this function can never be positive. The corresponding LP formulation for estimating such a function

using quantile regression is presented below.

$$\text{Minimize} \quad \sum_{n=1}^{N} Z_n \tag{B.1}$$

subject to:

$$Z_n \geq y_n - \left[\sum_i \alpha_k^i \theta_n^i + (\sum_i \beta_k^i \theta_n^i) x_n \right] \quad \text{if } k-1 \leq x_n \leq k, \forall n \tag{B.2}$$

$$Z_n \geq \omega_\tau \left[\sum_i \alpha_k^i \theta_n^i + (\sum_i \beta_k^i \theta_n^i) x_n - y_n \right] \quad \text{if } k-1 \leq x_n \leq k, \forall n \tag{B.3}$$

$$\sum_i \beta_k^i F_i(\theta_n^i) \leq 0, \ \forall k \in \{1, \cdots, x_{\max}\},$$
$$\forall F_i(.) \in \{\max_n(\cdot), \min_n(\cdot)\}, \ \forall i \tag{B.4}$$

$$\sum_i \alpha_k^i F_i(\theta_n^i) + [\sum_i \beta_k^i F_i(\theta_n^i)]k = \sum_i \alpha_{k+1}^i F_i(\theta_n^i) + [\sum_i \beta_{k+1}^i F_i(\theta_n^i)]k,$$
$$\forall k \in \{1, 2, ..., x_{\max} - 1\}, \ \forall F_i(.) \in \{\max_n(.), \min_n(.)\}, \ \forall i \tag{B.5}$$

The above formulation simply replaces the convexity-enforcing constraints (2.7) in Chapter 2 with the monotonic non-increments constraint (B.4).

Now, the goal is to compare the statistical fit achieved by the monotonic non-increment model for capacity envelope (M) with the more restricted convex model for capacity envelope (RM) using hypothesis testing. The task, similar to the discussion presented in Appendix I, reduces to identifying the degrees of freedom, in the form of linear equalities, that are arrested in unrestricted model (M) (monotonic non-increment) to obtain the restricted model (RM) (convex).

Given that the convexity constraint (2.7) is a restricted version of the monotonic non-increment constraints (B.4), we can split every estimated slope coefficient β_k^i of the monotonic non-increment model into two additive components $\beta_{k,\text{conv}}^i$ and $\beta_{k,\text{inc}}^i$, s.t. $\beta_k^i = \beta_{k,\text{conv}}^i + \beta_{k,\text{inc}}^i$ and the set of coefficient components $\beta_{k,\text{conv}}^i$ satisfy the convexity constraint. Now, the degrees of freedom separating the convex model estimates from the monotonic non-increment model estimates can be equated to the minimum number of incremental components $\beta_{k,\text{inc}}^i$ that need to be released to have a non-zero value in the above split-representation for the

monotonic non-increment estimates β_k^i. This can be determined through the integer program formulated below.

$$\text{Minimize} \quad \sum_{k=1}^{x_{\max}} \sum_i z_k^i \tag{B.6}$$

subject to:

$$\beta_k^i = \beta_{k,\text{conv}}^i + \beta_{k,\text{inc}}^i \qquad \forall k \in \{1,..,x_{\max}\}, \; \forall i \tag{B.7}$$

$$\beta_{1,\text{inc}}^i \leq 0 + Mz_1^i \qquad \forall i$$

$$\beta_{1,\text{inc}}^i \geq 0 - Mz_1^i \qquad \forall i \tag{B.8}$$

$$\beta_{k,\text{inc}}^i \leq \beta_{k+1,\text{inc}}^i + Mz_k^i \qquad \forall k \in \{1,..,x_{\max}-1\}, \; \forall i$$

$$\beta_{k,\text{inc}}^i \geq \beta_{k+1,\text{inc}}^i - Mz_k^i \qquad \forall k \in \{1,..,x_{\max}-1\}, \; \forall i \tag{B.9}$$

$$\sum_i \beta_{k+1,\text{conv}}^i F_i(\theta_n^i) \leq \sum_i \beta_{k,\text{conv}}^i F_i(\theta_n^i), \; \forall k \in \{1,\cdots,x_{\max}-1\},$$

$$\forall F_i(.) \in \{\max_n(\cdot), \min_n(\cdot)\}, \; \forall i \tag{B.10}$$

$$\sum_i \beta_{1,\text{conv}}^i F_i(\theta_n^i) \leq 0, \; \forall F_i(\cdot) \in \{\max_n(\cdot), \min_n(\cdot)\}, \; \forall i \tag{B.11}$$

$$z_k^i \in \{0,1\}, \quad \forall k \in \{1,..,x_{\max}\}, \; \forall i$$

Constraints (B.7), (B.10) and (B.11) enforce the above-described definition for the additive coefficient components $\beta_{k,\text{conv}}^i$ and $\beta_{k,\text{inc}}^i$. Constraint (B.9) ensures incremental slope components across adjacent linear pieces, $\beta_{k,\text{inc}}^i$ and $\beta_{k+1,\text{inc}}^i$, are equal if dummy binary variable $z_k^i = 0$, while constraint (B.8) ensures incremental component for the first linear piece $\beta_{1,\text{inc}}^i$ is 0 if $z_1^i = 0$. Note that these are similar to the set of constraints (A.2) in appendix I, which is used to identify the actual arrested degrees of freedom from (M) to (RM) in a piecewise-linear model framework. Under the above formulation, the optimal value of objective function (B.6) measures the minimum number of linear equalities of the set (B.9) and (B.8) that will have to be relaxed, by setting appropriate $z_k^i = 1$, to obtain the monotonic non-increment model estimates β_k^i from the convex model estimates $\beta_{k,\text{conv}}^i$. This in turn gives a measure of the linear degrees of freedom separating the convex and the monotonic non-increment model for capacity envelopes, and can therefore be used when performing the

Nested Hypothesis test (using likelihood ratios) for comparing the statistical fit of the above two models. This statistical test verifies the validity of the convexity assumption for capacity envelopes.

Appendix C

Determining order of quantile τ in capacity envelope estimation

The order of quantile τ in the capacity envelope functions expressed in (2.2) and (2.3) perform the same function as the specified frequency threshold in Gilbo (1993) [31] during empirical estimation of airport capacity. As depicted in Figure 2-3, this parameter eliminates outliers amongst the observed airport throughputs, typically points in the top $(1 - \tau/100)$%ile of the dataset, and ensures the resulting convex hull estimate is robust. Therefore, the choice of this parameter critically determines the quality of airport capacity estimate obtained from observed data. Gilbo (1993) [31] prescribes using frequency thresholds that correspond to a desired level of confidence in the resulting estimates. In this study, we adopt a similar principle related to the robustness of the estimates when choosing the order of quantile for the capacity functions. The basic idea is to identify the order of quantile τ for which the obtained estimates for the capacity envelope, including coefficients α_k^i and β_k^i, are stable. This is done through a trial-and-error process where τ is gradually lowered from 100%ile through small perturbations, till a relatively broad range of values for τ producing consistent values for estimates α_k^i and β_k^i is identified. Figure C-1 below, which presents the capacity estimates for different τ values on a sample dataset of throughputs, illustrates this process.

As observed in the figure, there is volatility in the capacity envelope estimates for τ values of 100 and 99.5 %ile, indicating the presence of outliers in this peripheral section of the planar scatter of observed throuphputs. The estimates are however stable and consistent
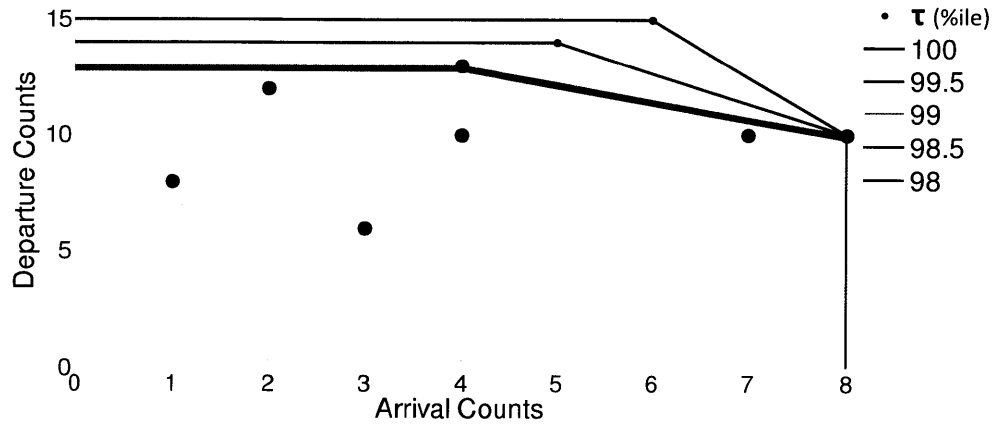
Figure C-1: Capacity envelope estimates for different values of τ

for τ values of 99, 98.5 and 98%ile, implying that this range of τ delivers a reliable empirical measure of the capacity envelope.

Appendix D

Determining flight-specific delay costs at LGA

This section describes the methodology used to obtain representative measures of unit ground and airborne delay costs for individual flights in the LGA case study discussed in Section 6.3.2. The data for this case study was extracted from a real-world GDP recorded on Feb 17, 2006 from 7am to midnight. The arrival demand over this 17-hr period comprised of 542 arrivals distributed across 27 operating airlines. The computed flight-specific delay costs were used for two purposes within the case study:

1. Calculate average unit ground and airborne delay costs (C_g and C_a) used in objective functions of the stochastic ground-holding models.

2. As input to the intra-airline slot substitution mechanisms executed on behalf of each airline in response to the stochastic ground-hold allocation.

Due to data limitations, the unit ground $C_{g,f}$ and airborne $C_{a,f}$ delay costs for flight f were assumed to be linear and made up of three main additive components that varied by aircraft type ac_f and operating airline ar_f: fuel, crew and passenger.

$$C_{g,f} = C_{g,\mathrm{ac}_f,\mathrm{ar}_f}^{\mathrm{fuel}} + C_{\mathrm{ac}_f,\mathrm{ar}_f}^{\mathrm{crew}} + C_{\mathrm{ac}_f,\mathrm{ar}_f}^{\mathrm{passenger}}$$
$$C_{a,f} = C_{a,\mathrm{ac}_f,\mathrm{ar}_f}^{\mathrm{fuel}} + C_{\mathrm{ac}_f,\mathrm{ar}_f}^{\mathrm{crew}} + C_{\mathrm{ac}_f,\mathrm{ar}_f}^{\mathrm{passenger}} \tag{D.1}$$

Note that the fuel cost is the only component distinguishing the unit delay cost incurred on the ground $(C_{g,f})$ from that on air $(C_{a,f})$ for any flight f. The ratio between ground and airborne fuel costs per unit time (ϕ_{ac}) was computed for prominent aircraft types using aircraft performance data reported in a study conducted under project GAES [69] on environmental impacts of air delay. Quarterly operating expenses of major airlines, as archived in P-5.2 tables of BTS database [72], provide fuel and crew costs incurred per block hour of flight operation for different aircraft types (i.e. $C_{a,\mathrm{ac},\mathrm{ar}}^{\mathrm{fuel}}$ and $C_{a,\mathrm{ac},\mathrm{ar}}^{\mathrm{crew}}$). These are used as approximate measures for fuel and crew components of unit delay costs of flights in the case study, with the fuel component of the unit ground delay cost calculated as $C_{g,\mathrm{ac},\mathrm{ar}}^{\mathrm{fuel}} = \phi_{\mathrm{ac}} \times C_{a,\mathrm{ac},\mathrm{ar}}^{\mathrm{fuel}}$.

The passenger delay component for unit flight delay costs are calculated using the following expression.

$$C_{\mathrm{ac}_f,\mathrm{ar}_f}^{\mathrm{passenger}} = 2 \times \mathrm{pass}_f \times \mathrm{pass_V OT} \tag{D.2}$$

where pass_f is the passenger volume for flight f, and $\mathrm{pass_V OT}$ is the passenger value of time. The multiplier of 2 is necessary to account for possibility of missed passenger connections resulting from flight delays. Barnhart et.al. (2008) [9] estimate the amount of passenger delays in NAS to be almost twice that of flight delays in their analysis, and this estimate is used as an average approximation of the delay per passenger resulting from unit flight delay in our cost computation.

The passenger volumes for flights pass_f in the case study are extracted from T-100 files in the BTS database [72] that contain airline-reported data on average load factors served in various domestic non-stop segments.

The official estimate of passenger value of time adopted by the FAA in their aviation policy planning [33] is used in this study as $\mathrm{pass_V OT}$.

Cost data for aircraft type, airline or flight segment that were missing from the above-described sources were filled in using appropriate interpolation.

Bibliography

[1] K. I. Ahmed. *Modeling drivers' acceleration and lane changing behaviors.* PhD thesis, Massachusetts Institute of Technology, 1999.

[2] K. I. Ahmed, M. Ben-Akiva, H. N. Koutsopoulos, and R. G. Mishalani. Models of freeway lane changing and gap acceptance behavior. In *Proceedings of the 13th International Symposium on the Theory of Traffic Flow and Transportation*, 1996.

[3] M. O. Ball, R. Hoffman, D. Lovell, and A. Mukherjee. Response mechanisms for dynamic air traffic flow management. In *Proc. 6th USA/Europe Air Traffic Management R & D Seminar, Baltimore, MD*, December 2005.

[4] M. O. Ball, R.Hoffman, A.R.Odoni, and R.Rifkin. A stochastic integer program with dual network structure and its application to the ground-holding problem. *Operations Research*, 51(1):167–171, 2003.

[5] M. O. Ball, R.Hoffman, D. Knorr, J. Wetherly, and M. Wambsganss. Assessing the benefits of collaborative decision making in air traffic management. In *Proc. of 3rd USA/Europe Air Traffic Management R & D Seminar*, December 2005.

[6] M. O. Ball, R.Hoffman, and A. Mukherjee. Ground delay program planning under uncertainty based on the ration-by-distance principle. *Transportation Science*, 44(1):1–14, 2010.

[7] Michael Ball, Cynthia Barnhart, Martin Dresner, Mark Hansen, Kevin Neels, Amedeo Odoni, Everett Peterson, Lance Sherry, Antonio Trani, and Bo Zou. Total Delay Impact Study: A Comprehensive Assessment of the Costs and Impacts of Flight Delay

in the United States. Technical report, National Center for Excellence For Aviations Operations Research, 2010.

[8] C. Barnhart, D. Bertsimas, C. Caramanis, and D. Fearing. Equitable and efficient coordination of traffic flow management programs. *Working paper*, 2009.

[9] C. Barnhart, D. Fearing, and V. Vaze. Modeling passenger travel and delays in the national air transportation system. under review. *Operations Research*, 2011.

[10] Moshe Ben-Akiva and Steve Lerman. *Discrete choice analysis: theory and application to travel demand.* MIT Press, 1985.

[11] C. Bernini, M. Freo, and A. Gardini. Quantile estimation of frontier production function. *Empirical Economics*, 29:373–381, 2004.

[12] D. Bertsimas and S. Gupta. Fairness in air traffic flow management. Technical report, Massachusetts Institute of Technology, 2009.

[13] D. Bertsimas and S. Stock Patterson. The air traffic flow management problem with enroute capacities. *Operations Research*, 46:406–422, 1998.

[14] Michel Bierlaire. Biogeme: a free package for the estimation of discrete choice models. In *3rd Swiss Transport Research Conference*, 2003.

[15] A. Blumstein. The landing capacity of a runway. *Operations Research*, 7(1):752–763, 1959.

[16] Philippe A. Bonnefoy. Emergence of Secondary Airports and Dynamics of Regional Airport Systems in the United States. Master's thesis, Massachusetts Institute of Technology, 2005.

[17] Bureau of Transportation Statistics. BTS website, 2007. http://www.bts.gov.

[18] K. Chang, K. Howard, R. Oisen, L. Shisler, M. Tanino, and M. Wambsganss. Enhancements to the faa ground delay programs under collaborative decision making. *Interfaces*, 31:57–76, 2001.

[19] D. R. Cox. Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society. Series B (Methodological)*, 24(2):406–424, 1962.

[20] R. J. Dakin. A tree search algorithm for mixed integer programming problems. *Computer Journal*, 8(3):250–255, 1965.

[21] Richard de Neufville and Amedeo Odoni. *Airport Systems: Planning, Design and Management*. McGraw-Hill, 2003.

[22] D.R.Cox. Tests of separate families of hypotheses. *Proceedings of 4th Berkeley Symposium*, 4(1):105–123, 1961.

[23] Federal Aviation Administration. FAA CDM website, 2006. `http://cdm.fly.faa.gov`.

[24] Federal Aviation Administration. OPSNET database, 2007. `http://aspm.faa.gov/main/opsnet.asp`.

[25] Federal Aviation Administration. Aviation System Performance Metrics, accessed September 2008. `http://aspm.faa.gov`.

[26] R. Fourer, D. M. Gay, and B. W. Kernighan. *AMPL: A Modeling Language for Mathematical Programming*. The Scientific Press (now part of Wadsworth Publishing), 1993.

[27] M. J. Frankovich, D. Bertsimas, and A. R. Odoni. Optimal selection of airport runway configurations. In *INFORMS Annual Meeting, San Diego, CA*, 2009.

[28] M. Furno. Parameter instability in quantile regression. *Statistical Modelling*, 7(4):345–362, 2007.

[29] E. Gilbo and K. Howard. Collaborative optimization of airport arrival and departure traffic flow management strategies for CDM. In *Proceedings of 3rd USA/Europe Air Traffic Management R&D Seminar*, 2000.

[30] E. P. Gilbo and K. W. Howard. Collaborative optimization of airport arrival and departure traffic flow management strategies for cdm. In *3rd USA/Europe Air Traffic Management R & D Seminar*, June 2000.

[31] E.P. Gilbo. Airport capacity: representation, estimation, optimization. *IEEE Transactions on Control System Technology*, 1(3):144–153, 1993.

[32] E.P. Gilbo. Optimizing airport capacity utilization in air traffic flow management subject to constraints at arrival and departure fixes. *IEEE Transactions on Control System Technology*, 5(5):490–503, 1997.

[33] GRA, Incorporated. Economic Values for FAA Investment and Regulatory Decisions, A Guide. Technical report, 2004.

[34] Air Transport Action Group. The economic and social benefits of air transport 2008. Technical report, 2008.

[35] William D. Hall. *Efficient Capacity Allocation in a Collaborative Air Transportation System*. PhD thesis, Massachusetts Institute of Technology, 1999.

[36] Michael J. Hanowsky. *A Model to Design a Stochastic and Dynamic Ground Delay Program Subject to Non-Linear Cost Functions*. PhD thesis, Massachusetts Institute of Technology, 2008.

[37] R.M. Harris. Models for runway capacity analysis. Technical report, The MITRE Corporation, MTR-4102, Rev. 2, Langley, Virginia, USA, 1972.

[38] J. Hausman and D. McFadden. Specification tests for the multinomial logit model. *Econometrica*, 52(5):1219–1240, 1984.

[39] H.M.Wagner. *Principles of Operations Research*. Prentice Hall, Englewood Cliffs, NJ, 1975.

[40] S.L.M. Hockaday and A. Kanafani. Development in airport capacity analysis. *Transportation Research*, 8(3):171–180, 1974.

[41] K. Howard. *ETMS/ATMS system requirements: slot credit substitutions (SCS) version 1.2*. John A. Volpe Res. Ctr., Cambridge, MA, 2002.

[42] ILOG, Inc. ILOG CPLEX 11.0.1, 2008. http://www.ilog.com.

[43] M. Janic and V. Tosic. Terminal airspace capacity model. *Transportation Research Part A*, 16(4):253–260, 1982.

[44] Joint Planning and Development Office. Next Generation Air Transportation System Integrated Plan, December 2004. `http://www.jpdo.aero/integrated_plan.html`.

[45] Mi-Ok Kim. Quantile regression with varying coefficients. *The Annals of Statistics*, 35(1):92–108, 2007.

[46] R. Koenker and K.F. Hallock. Quantile regression. *Journal of Economic Perspectives*, 15(4):143–156, 2001.

[47] Roger Koenker. *Quantile Regression*. Cambridge University Press, 2005.

[48] B. Kotnyek and O. Richetta. Equitable models for the stochastic ground-holding problem under collaborative decision making. *Transportation Science*, 40(2):133–146, 2006.

[49] B. Kotnyek and O. Richetta. Equitable models for the stochastic ground-holding problem under collaborative decision making. *Transportation Science*, 40(2):133–146, 2006.

[50] D.A. Lee, P.F. Kostiuk, R.V. Hemm, W.E. Wingrove, and G. Shapiro. Estimating the effects of the terminal area productivity program. Technical report, Logistics Management Institute, NS301R3, 1997.

[51] Leihong Li and John-Paul Clarke. A stochastic model of runway configuration planning. In *AIAA Guidance, Navigation and Control Conference*, August 2010.

[52] G. Lulli and A. R. Odoni. The european air traffic flow management problem. *Transportation Science*, 41(4):431–443, 2007.

[53] A. R. Odoni M. A. Stamatopoulos, K. G. Zografos. A decision support system for airport strategic planning. *Transportation Research Part C*, 12:91–117, 2004.

[54] Mitre Corporation and Federal Aviation Administration. Airport capacity benchmark report 2004. Technical report, 2004.

[55] A. Mukherjee and M. Hansen. A dynamic stochastic model for the single airport ground holding problem. *Transportation Science*, 41(4):444 456, 2007.

[56] A. Mukherjee, M. Hansen, and S. Grabbe. Ground delay program planning under uncertainty in airport capacity. In *AIAA Guidance, Navigation and Control Conference*, August 2009.

[57] Avijit Mukherjee. *Dynamic Stochastic Optimization Models for Air Traffic Flow Management*. PhD thesis, University of California, Berkeley, 2004.

[58] G.F. Newell. Airport capacity and delays. *Transportation Science*, 13(3):201–241, 1979.

[59] A. R. Odoni, L. Bianco, and G. Szego. The flow management problem in air traffic control. *Flow Control of Congested Networks, Spinger-Verlag, Berlin*, pages 269–288, 1987.

[60] A. R. Odoni, J. Bowman, D. Delahaye, J.J. Deyst, E. Feron, R.J. Hansman, K. Khan, J.K. Kuchar, N. Pujet, and W.R. Simpson. Existing and required modeling capabilities for evaluating ATM systems and concepts. Technical report, International Center for Air Transportation, Massachusetts Institute of Technology, NAG2-997, 1997.

[61] Michael Downes Peterson. *Models and Algorithms for Trabsient Queueing Congestion in Airline Hub and Spoke Networks*. PhD thesis, Massachusetts Institute of Technology, 1992.

[62] Robert S. Pindyck and Daneil L. Rubinfeld. *Econometric Methods and Econometric Forecasts*. McGraw-Hill, 1998.

[63] P.Kall. *Stochastic Linear Programming*. Springer, Berlin, 1976.

[64] Nikolas Pyrgiotis. *A Stochastic and Dynamic Model of Delay Propagation Within an Airport Network for Policy Analysis*. PhD thesis, Massachusetts Institute of Technology, 2011.

[65] V. Ramanujam and H. Balakrishnan. Estimation of arrival-departure capacity tradeoffs in multi-airport systems. In *IEEE Conference on Decision and Control*, December 2009.

[66] V. Ramanujam and H. Balakrishnan. Estimation of discrete choice models of the runway configuration selection process. In *American Control Conference*, June 2011.

[67] O. Richetta and A. R. Odoni. Solving optimally the static ground holding policy problem in air traffic control. *Transportation Science*, 27(3):228–238, 1993.

[68] O. Richetta and A. R. Odoni. Dynamic solution to the ground-holding problem in air traffic control. *Transportation Research Part A*, 28:167–185, 1994.

[69] Sandrine Carlier and Jean-Claude Hustache (ENVISA), Frank Jelinek (EUROCONTROL). Project GAES: Environmental Impact of Delay. Technical report, 2006.

[70] I. Simaiakis, A. Donaldson, and H. Balakrishnan. Impact of heavy aircraft operations on airport capacity at ewr. In *AIAA Aviation Technology, Integration, and Operations Conference*, September 2011.

[71] W.J. Swedish. Upgraded airfield capacity model. Technical report, The MITRE Corporation, MTR-81W16, 1981.

[72] U.S. Department of Transportation. Bureau of Transportation Statistics, accessed September 2011. http://www.transtats.bts.gov.

[73] C. S. Venkatakrishnan, A. Barnett, and A. R. Odoni. Landings at Logan airport: Describing and increasing airport capacity. *Transportation Science*, 27(3):211–227, 1993.

[74] P. Vranas, D. Bertsimas, and A. R. Odoni. The multi-airport ground holding problem for air traffic control. *Operations Research*, 42:249–261, 1994.

[75] M. Wambsganss. Collaborative decision making through dynamic information transfer. *Air Traffic Control Quarterly*, 4:107–123, 1996.