ABSTRACT

Title of Document: Optimization Models for Speed Control in Air Traffic Management

James C. Jones, Doctor of Philosophy, 2015

Directed By: David J. Lovell, Associate Professor, Department of Civil & Environmental Engineering

In a typical air traffic control environment, the precise landing times of en route aircraft are not set until each aircraft approaches the airspace adjacent to the destination airport. In times of congestion, it is not unusual for air traffic controllers to subject arriving aircraft to various maneuvers to create an orderly flow of flights onto an arrival runway. Typical maneuvers include flying in zig-zag patterns, flying in circular holding patterns and tromboning. These maneuvers serve to delay the arrival time of the flight while also burning additional fuel. On the other hand, if the arrival time was established much earlier, then such delay could be realized by simply having flights fly slower while still at a higher altitude, which would incur much less fuel burn than the described maneuvers. Yet despite its potential benefit, thus far little has been done to promote the management of flights using speed control in the presence of uncertainty.

This dissertation presents a set of models and prescriptions designed to use the mechanism of speed control to enhance the level of coordination used by FAA managers at the tactical and pre-tactical level to better account for the underlying uncertainty at the time of planning. Its models deal with the challenge of assigning delay to aircraft approaching a single airport, well in advance of each aircraft's entry into the terminal airspace. In the first approach, we assume control of all airborne flights at a distance of 500 nm while assuming no control over flights originating less than 500 nm from the airport. We propose a set of integer programming models designed to issue arrival times for controlled flights in the presence of the uncertainty associated with the unmanaged flights. In the second approach, we assume control over all flights by subjecting flights to a combination of air and ground delay. Both approaches show strong potential to transfer delay from the terminal to the en route phase of flight and achieve fuel savings. Building on these ideas we then formulate an approach to incorporate speed control into Ground Delay Programs. We propose enhancements for equitably rationing airport access to carriers and develop a revised framework to allow carriers to engage in Collaborative Decision Making. We present new GDP control procedures and also new flight operator GDP planning models. While the ability to achieve all the benefits we describe will require NextGen capabilities, substantial performance improvements could be obtained even with a near-term implementation.

OPTIMIZATION MODELS FOR SPEED CONTROL IN AIR TRAFFIC MANAGEMENT

By

James C. Jones Jr.

Dissertation submitted to the Faculty of the Graduate School of the University of Maryland, College Park, in partial fulfillment of the requirements for the degree of Doctor of Philosophy
2015

Advisory Committee:
Professor David J. Lovell, Chair
Professor Michael O. Ball
Professor Steven Gabriel
Professor John-Paul Clarke
Dr. Robert Hoffman

Dedication

To my Mother, Father and Grandmother.

Acknowledgements

So here we are, at the beginning but at the end. It has been a great privilege for me to be involved in the world of air transportation research during my time at Maryland and there are many to thank for helping to make it happen. I would like to thank my advisor David Lovell for his invaluable advice, mentorship and feedback which has served to make me a better researcher. I would also like to thank Michael Ball for introducing me to integer programming and providing thoughtful and instrumental guidance in my research. I am very appreciative of them both for helping to get me in the game.

I would like to thank my other committee members John-Paul Clarke, Bob Hoffman and Steven Gabriel. It was a pleasure to work with John-Paul and his students during my first two years, and our discussions during this time have been a real asset to my research. I would like to thank Bob for providing a more practical viewpoint for our problems and his helpful suggestions over the evolution of this effort. I am also grateful to Steven Gabriel for his assistance on this dissertation and for providing such excellent instruction in mathematical programming.

In addition to those already mentioned, my perspectives on speed control and air transportation have also been shaped by discussions with Luis Delgado, David Rey, Bosung Kim, Jim Brooks and Frank Ketcham. I am grateful to each of them for taking the time to lend a critical ear and an insightful perspective.

During my time in school I had a great experience interning at NASA Ames. I am very appreciative of Banavar Shridar, Alan Lee, Steve Green and Gilbert Wu for helping to make it happen. I really enjoyed collaborating with Gilbert and Steve and getting to know many of the other NASA researchers.

It has been my pleasure to share office space, ideas and good times with Kleoniki Vlachou, Carina Wang, Prem Swaroop, Xenia Barnes, Alex Estes and Max Hoffman. I owe great thanks to Brian Waskiewicz for lending his operational knowledge as a pilot on many an occasion; he has been a great sounding board and a better friend. I would also like to thank my friends from Swarthmore, Maryland, Los Angeles and elsewhere for their support and comradery over the years. To Rebecca, I am grateful for the love and support she has shown me and her willingness to listen to my ideas about air transportation and anything else on my mind.

The fact that I have gotten this far is a testament to the sacrifices my parents Jim and Colleen Jones made to help me succeed for which I am truly grateful. I am thankful for their help in getting me started and keeping me on the right track through those formative years. I am also appreciative of my grandmother Carolyn for maintaining a strong interest in my work over the years. The fact that you still take time to read my papers and ask me questions is truly inspiring. I hope I have been a good reflection on you.

Table of Contents

List of Tables

List of Figures

1 Introduction

In the United States, air carriers are free to design and operate their flight schedule at whatever times best suit their own business objectives. While this policy works well in many instances, occasionally the demand for certain resources within the National Airspace System (NAS) well exceeds the capacity of the system to accommodate it. For example, bad weather can severely impact the capacity of an airport to land arriving aircraft. These imbalances can be grouped into two categories: those that occur at the pre-tactical level and those occurring at the tactical level. At the pre-tactical level, congestion is forecast hours in advance of its occurrence and in these situations, the Federal Aviation Administration (FAA) develops a plan to mitigate the potential congestion in the affected area. At the tactical level, such imbalances may only be known minutes before they materialize. Here, air traffic controllers reactively modify the trajectories and speeds of the incident flights to ensure safety is maintained within the system.

Despite the differences in dealing with the two situations, they are often intertwined. Severe weather patterns will often emerge near airports and the surrounding regions, causing the FAA to impose some sort of advance delay. Yet, due to uncertainty in departure times, turbulent weather en route and runway availability, flights may also deviate from their planned arrival times imposed by the FAA. When this happens, air traffic controllers are left to handle the additional delay. The lack of coordination between parties can cause the affected flights to burn additional fuel due to airborne holding. Previous research has not addressed the issue of coordination between stakeholders at the pre-tactical and tactical level, nor its effect on the workload of the air traffic controllers.

This dissertation presents a set of models and prescriptions designed to use the mechanism of speed control to enhance the level of coordination used by FAA managers at the pre-tactical level to better account for the underlying uncertainty at the time of planning. While spatial deviations and circular holding patterns are well-known tools used to occupy aircraft queued up for a congested airport, the idea of speed control is much less common, and requires significantly more advanced planning. The basic idea is that, armed with good forecasts of the situation at the airport, the speeds, and hence arrival times, of en route flights can be planned systematically to best condition the arriving traffic stream for smooth integration with the airport operation.

In the next section, we describe in more detail programs and mechanisms currently used to issue issued planned delays to the affected flights. Subsequently, we present the set of models and systems that have been used to date to provide better air traffic management and speed control guidelines to flights.

1.1 Air Traffic Management

Modern Air Traffic Management (ATM) consists of two components: traffic flow management and air traffic control. Air traffic control functions at the tactical level and affects flights over a period of seconds to 30 minutes. Air traffic control aims to ensure safe separation between aircraft through rerouting, speed control, navigational vectoring and airborne holding. The traffic flow management component operates at the strategic level where flights are managed over a horizon of 30 minutes to 19 hours and are sometimes planned days in advance. It seeks to ensure the smooth flow of aircraft through airspace by balancing the demand for air traffic with the capacity of the available air traffic resources (Ball, et al., 2007; De Neufville & Odoni, 2003). In this section, we describe the relevant mechanisms used to facilitate traffic flow management and highlight the challenges with incorporating some the underlying principles in the Next Generation Air Traffic Management System (NextGen).

1.1.1 The Need for Coordinated Planning

In recent years, heightened levels of congestion at major airports across the U.S. have cost carriers billions of dollars in lost revenue. By one estimate, these delays cost the airline industry $8.178 billion in 2012. The extent of such a loss can be quite severe considering that in the same year, the top 10 carriers only made a profit of $5.31 billion (U. S. Department of Transportation, 2015). The impact of these delays, however, is not solely felt by the airlines, as the effect on passengers in some years can be just as much. In 2007, it is estimated that flight delays cost passengers an estimated $16.7 billion in lost time in a year where airlines delay costs totaled $8.3 billion (Ball, M.; Barnhart, C.; Dresner, M.; Hansen, M.; Neels, K.; Odoni, A.; Peterson, E.; Sherry, L.; Trani, A., Zou, B., 2010).

In 2013, 16.65% of all flights in the United States were delayed by 15 minutes or more (Bureau of Transportation Statistics, 2015). Of these delays, weather accounted for 58.69% of the delay, while volume accounted for 28.3%. Additionally, 1.29% of all flights were cancelled. In both cases the delay is caused by persistent mismatches between capacity and demand (Airlines for America, 2015). The problem will likely be exacerbated if demand continues to increase, as its effect on delay is nonlinear. A 1% increase in demand can often lead to a 10% increase in delay when an airport is operating at or near capacity (Odoni, 2009).

There are a number of prescriptions for dealing with rising demand, and almost all fall into one of three types of approach. The first type of approach seeks to increase the physical capacity by creating more airports and runways. This is quite a difficult undertaking, as it requires considerable time, funding and buy-in from government, industry and community stakeholders. The second approach attempts to manage demand by imposing administrative and/or economic measures on carriers to convince them to reduce their demand or shift it to less congested periods of time. Such mechanisms include schedule coordination, congestion pricing, slot auctions and slot trading. Irrespective of the success of these long-term measures, the NAS will continue to be affected by periodic perturbations arising from inclement weather and runway closings. Air traffic flow management offers a means to deal with these eventualities on a more granular level; the FAA has developed a number of Traffic Management Initiatives to coordinate traffic in affected areas to better match capacity to demand. These tools have proven essential in ensuring continued functionality within the NAS.

1.1.2 Traffic Management Initiatives

The Air Traffic Control System Command Center (ATCSCC) in Warrenton, VA monitors Air Traffic Flow Management (ATFM) operations on an on-going basis. Projections are periodically made to forecast the delays in the airspace and at airport resources throughout the country. When the projected demand for a resource exceeds its capacity by a sufficient amount, the ATCSCC uses a traffic management initiative (TMI) to manage the demand. These strategic-level TMIs include ground delay programs (GDPs), airspace flow programs (AFPs) and ground stop programs. Less programmatic measures also include Miles-in-Trail (MIT) restrictions, rerouting and airborne holding. These initiatives are described below.

- *Ground Delay Program:* When the number of flights scheduled to arrive at an airport over a sustained period of time exceeds its forecast capacity, the flights scheduled to arrive at that airport are delayed on the ground at their origin airport(s) prior to departure, with assigned delays cascading according to their scheduled arrival times at the destination airport. The resulting flight delay reduces the rate of flight arrivals to a level that the affected airport can accommodate. Flights are given a controlled time of departure (CTD) that is often later, but certainly no earlier, than their scheduled departure times, based on the order that they appear in the flight schedule. The rationale behind the program lies in the fact that it is cheaper and safer for the airline and the system to absorb a flight delay on the ground than by holding it in the air. The duration of a GDP is typically a few hours; however, programs occasionally exceed 19 hours. Carriers occasionally prefer to cancel flights rather than absorb these delays; presumably this is done in cases where those passengers could easily be accommodated on other flights, and the airframe in question is not needed downstream.
- *Airspace Flow Program:* When the capacity of the airspace is insufficient to deal with the predicted number of flights that will fly through it, the capacity

is rationed. As with GDPs, flights are delayed on the ground to stem the flow of traffic into the affected airspace. Flights also receive a CTD that is reflective of their order in the flight schedule. Again, carriers can cancel flights to avoid this, but they can also re-route their flights around the bad weather (with permission from the FAA), which is not commonly done during ground delay programs.

- *Ground Stop Program:* When the capacity of an airport is diminished to the point that it is unable to accommodate incoming flights, departing flights are delayed on the ground until the affected airport can again begin to accept new flights. Additionally ground stops are also initiated to allow time for the implementation of a more long-term solution such as a GDP.
- *Airborne Holding:* When a sector or airport is unable to accommodate the number of incoming flights and the level of demand was not sufficiently reduced on the ground, controllers will often hold flights in the air until they can be accommodated. This type of delay is generally undesirable because it is more expensive for the airlines to incur and it imposes an additional burden on air traffic controllers.
- *Miles-in-Trail (MIT) Restrictions:* Another means of controlling sector and airport demand is to delay flights as they move en route. MIT restrictions impose spacing limits on succeeding aircraft to reduce the rate of arrival at the affected resource. In effect, MIT restrictions impose delay on flights by slowing the aircraft. While such restrictions are typically less efficient than time-based speed control initiatives, the resulting delay produced is generally

less expensive than airborne holding as flights do not have to travel significant additional distances to acquire the delay, and they can incur the delays at higher altitudes where fuel consumption is lower.

- *Rerouting:* When a sector or region of airspace encounters bad weather, it is often desirable to allow flights to move along an adjusted trajectory where it can more easily be managed. In these situations, flights are routed onto alternative routes through areas that are not affected by weather. This rerouting can produce additional congestion in the areas of routing due to the additional traffic. Thus, these areas often require MIT restrictions in order to ease the ensuing congestion.

1.1.3 Collaborative Decision Making

The decision-making duties involved with managing ATFM operations are performed by the both the airlines and the FAA. While the FAA can impose oversight to ensure the safe operation of NAS resources, it does not decide whether to schedule or cancel flights. As such, effective coordination of TMI requires input from both sets of stakeholders. For example, the FAA could impose a GDP at Boston Logan Airport due to severe weather only to find out that many of the airlines had decided to cancel a substantial number of their flights. These cancellations might have reduced the demand to a level that the airport could accommodate; however, the carrier does not have any incentive to inform the FAA that those cancellations were made (thereby freeing up capacity for its competitors). As a result, without some collaboration mechanism, lack of shared information can cause inefficient use of constrained resources, and thereby produce significant delays. To enhance the level of cooperation and information sharing between stakeholders, the FAA adopted a philosophy known as Collaborative Decision Making (CDM) at beginning of the 21st century (Ball, et al., 2007; Wambsganss, 1996). This philosophy put more of the decision-making responsibilities in the hands of the carriers instead of the FAA. By allowing carriers to participate more actively in the process, CDM allows airlines to better determine how the delays get allocated across the flights that they operate. The goals of CDM are summarized below; for a more detailed description see Ball et al, (2007).

- Generate better information by combining the data collected through airspace monitoring with flight data for carriers.
- Ensure common situational awareness by disseminating shared information between the FAA and carriers.
- Provide tools and procedures that allow carriers to respond to capacity/demand imbalances while working in concert with the FAA to prescribe flow management of aircraft.

The CDM resource allocation mechanism for GDP planning consists of three components: capacity allocation, schedule adjustments and slot exchange. While control is executed based on a CTD, planning is done based on a controlled time of arrival (CTA). Specifically, arrival capacity is allocated to carriers using a mechanism known as Ration-by-Schedule (RBS). To improve throughput, an inter-airline substitution procedure known as compression is used facilitate trades. A notional diagram of the process is shown in Figure 1.1.

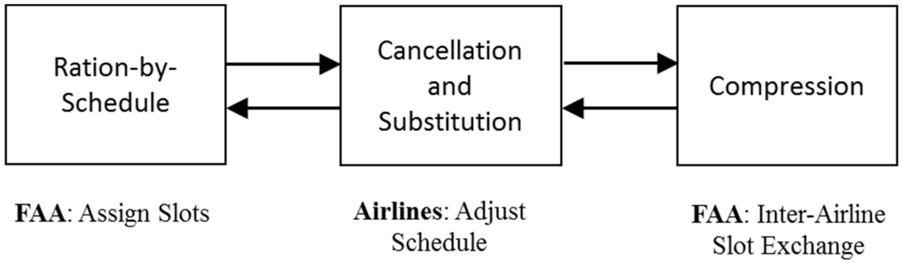

Figure 1.1 Flight Assignment during CDM

Ration-by-schedule (RBS) served as an initial mechanism for the FAA to assign capacity to airlines. Delayed flights are assigned arrival times based on the order that they appear in the schedule but no earlier than their scheduled time of arrival (STA). An example of the RBS procedure is shown in Figure 1.2. This assignment procedure, known as RBS, has become widely accepted by stakeholders as a standard for equitable allocation (Vossen & Ball, 2006a; Vossen & Ball, 2006b). The presumption is that the scheduled order was acceptable to the carriers, because they determined it when they scheduled their flights. Under a ground delay program, although the arrival times cannot be maintained, the order can, and this is deemed an equitable starting point.

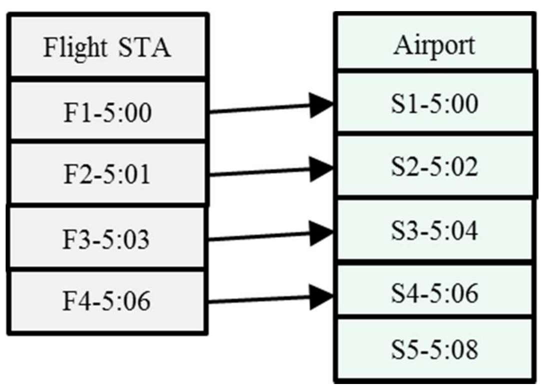

Figure 1.2 An example of an RBS allocation.

Once the capacity has been initially allocated through RBS, the carriers are free to substitute their allocation on an intra-airline basis. They may also cancel flights that they no longer wish to operate. After these substitutions and cancellations have been made, and announced to the FAA, the FAA runs a process of inter-airline slot trading known as compression. Compression fills the open slots that were made available by airline cancellations by allowing flights to move up in the queue. By allowing carriers to substitute their own flights whenever possible before giving other airlines access to the slots, the process rewards carriers for reporting cancellations. An illustration of the compression algorithm is shown in Figure 1.3.

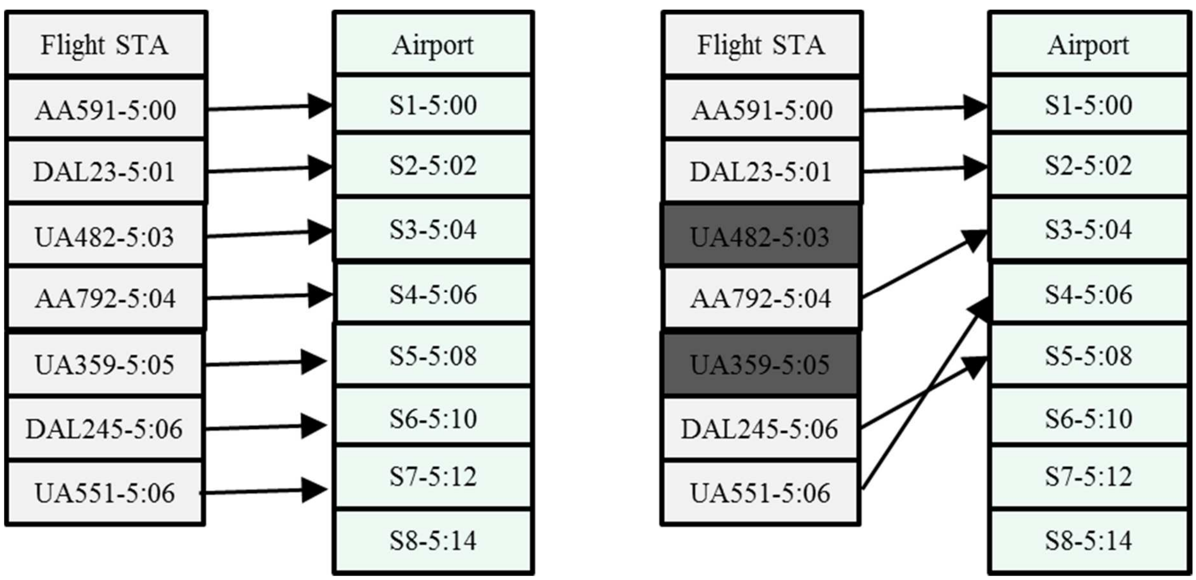

Figure 1.3 An example of the compression algorithm.

In this example, United Airlines has cancelled two flights after they received an allocation through RBS. The FAA is then able to fill these open slots with flights that would have otherwise received longer delays. Note that flight UA551 is able to receive the slot that was previously allocated to UA359 because the airline "owned" the slot. It cannot however, receive the slot that belonging to UA482 because it cannot receive a slot earlier than its STA.

When used in the context of CDM, RBS has several significant properties that have led to its adoption and popularity in practice:

1. In a single-resource setting, it minimizes the total delay (Ball, et al., 2007).
2. It lexicographically minimizes delay thereby ensuring standard of fairness that has been widely accepted by stakeholders (Ball, et al., 2007). Meaning that if A is a vector with the distribution of flight delays, N is the maximum number of minutes of delay assigned to any flight, and a_j is the set of flights receiving j minutes of delay, then for j = 0, 1, 2, …, N, RBS will order the flights as $A=(a_N, \ldots, a_1, a_0)$.
3. It avoids assigning a "double penalty" to airlines that are delayed due to other factors such as aircraft maintenance and crew availability. Thus it encourages accurate reporting of information (Vossen & Ball, 2006a).

There are, however, a few flaws associated with the RBS practice that have been identified over the course of its adoption. In a multi-resource setting, RBS does not necessarily yield the most efficient allocation of slots (Lulli & Odoni, 2007). If the affecting weather clears before the end of the planned GDP horizon, flights that are further away from the airport may be needlessly delayed (Ball, et al., 2010). To deal with this issue, the FAA allows flights beyond a specified radius to be exempt from the GDP. These exempted flights may arrive at whatever time best suits the needs of their operators. These exemptions can compromise the notion of equitability of the procedure. In some cases, the exemptions can produce biases that favor certain carriers over others (Vossen, et al., 2003). In Chapter 4, we discuss some methods for curbing this exemption bias using en route speed control and later propose removing the radius entirely to facilitate more airline-centric rescheduling through cancellation and substitution procedures. In practice, the setting of the exemption radius is a subjective exercise conducted during a strategic telecon between the FAA, carriers, and others. It is not done systematically, predictably, or objectively. As a result, some have suggested alternative methods of determining TMI parameters Swaroop and Ball (2013); Evans et al., (2014); Liu and Hansen (2013).

In AFPs the picture is a bit more complex. In these TMIs, traffic managers create a Flow Constrained Area (FCA) to reduce flow through the airspace. Once the FCA is defined to a set space and time flights that are scheduled to pass through it are assigned CTDs using RBS. The operators of these flights can then take the ground delays as given, substitute and cancel flight based on their internal priorities and/or reroute certain flights around the FCA. One issue with this process is that carriers can only assert their preference after the FCA has been defined by the FAA. This can lead to significant inconsistencies between what the allocated capacity and interests of carriers. To provide more opportunities for rerouting the FAA has developed the Collaborative Trajectory Options Program (CTOP). The initiative will introduce a few significant changes to the current set of practices: (Vlachou, 2014)

- Allow carriers to express their preferences for flight reroutes before the FCA is defined
- Allow the FAA and carrier to communicate electronically

- Use algorithms and automation in decision making to limit the impact of human error.

The system has been implemented by the FAA, however, carriers have yet to make the necessary changes to comply with the new framework. Once programs have fully adopted the changes will aim to provide carriers with better planning during the initial stages of the AFP and better communication and situational awareness throughout the programs.

1.1.4 Trajectory Based Operations and Collaborative Management

In recent years, much of the development in air traffic management has been oriented towards revamping the national air traffic management system, in order to produce a new technological paradigm, often referred to as NextGen. While a number of improvements have fallen under the guise of its development, the goals of the project include developing a platform to facilitate collaborative capacity and flow management, efficient trajectory management and flexible separation management. Powered by the additions of global position systems (GPS), onboard Flight Management Systems (FMS), enhanced navigational capability and better communication of information through datalink, the innovations of NextGen seek to improve access to airspace while increasing safety, capacity, efficiency and sustainability (Joint Planning and Development Office, 2011).

Leveraging the success of CDM in GDPs/AFPs, collaborative capacity and trajectory management attempts to foster an environment of greater shared situational awareness of available and active NAS routes and resources. In this paradigm, operators would share information with the FAA as they plan their flights. The FAA can then provide feedback about the viability of those plans. When plans are revised, the FAA managers and controller are updated with new information. This increased information sharing allows carriers to execute more control and flexibility over their flights while increasing the capacity of airspace available to flight operators.

In addition to improved collaborative decision-making, NextGen aims to provide greater involvement for ATC through a mechanism called Trajectory Based Operations (TBO). In this framework, flights would fly negotiated 4D trajectories, which allow for more precise active management across all phases of flight. Unlike the current environment, which relies heavily on voice communication, information exchange will take place increasingly over datalink communication. The negotiated trajectories will reduce the likelihood of conflict between flights, thereby increasing safety and reducing residual delay and fuel burn. In order to facilitate this NextGen development, the ATM community needs to define new ways to implement TMI over shorter time frames with ground delays, speed control and rerouting.

1.2 Literature Review

1.2.1 The Use of Speed Control

Air traffic flow management represents a problem domain where implicit or explicit modeling of uncertainty is crucial due to the strong roles played by fluctuations in weather, the human elements involved in air traffic control, mechanical delays, delays in aircraft turnaround, crew arrival and the complexity of airport surface operations. Thus, it is critical for decision support tools and models to implicitly or explicitly take uncertainty into account in producing recommended actions. Until recently in the U.S., there was no operational coordination of the arrival times of flights until the traffic management advisor (TMA) system exercises control in the general vicinity of the airport (starting approximately 200 nm out) (Swenson, et al., 1997). Additionally, the FAA has begun to deploy its Time-Based Flow Metering system. The system assigns speed advisories to flights along their trajectories every 200 nm and sets a freeze horizon inside of which the assigned arrival times are fixed. While these systems impose a high degree of control on flights arriving at these metering fixes they do not control flights originating inside the freeze horizon. This can impose considerable uncertainty on the system (Miller, et al., 2014). Even under departure controls such as Ground Delay Programs (GDPs), while departure times may be fairly well regulated, research has shown that there can still be considerable uncertainty in the arrival times of the aircraft, as measured against what the program expected them to do (Ball, et al., 2001). As a result, flights set their own speed profiles, and can even accelerate to attempt to make their scheduled arrival times, only to be subjected to maneuvers in terminal airspace to temporarily stem the flow of traffic into the destination airports, wasting a considerable amount of time and fuel. In this dissertation we present stochastic optimization models to determine delay transfer strategies. The long-term vision for air traffic management in both the U.S. (NextGen) and Europe (SESAR) calls for a move to trajectory based operations (TBO) under which trajectory timing would be set well in advance, leading, in concept, to a solution to the problem of excessive terminal area delays. However, the full implementation of TBO remains 15 to 20 years in the future and there are many research questions that must be answered in order to accomplish it. Our models can be viewed as a solution to this problem that could be implemented in the near-term or, alternatively, as an initial step toward TBO implementation.

The use of en route speed adjustments to achieve fuel savings and throughput benefits has been studied for over two decades. Neuman and Erzberger (1991) present a number of sequencing and spacing algorithms designed to reduce fuel consumption and en route/arrival delay. Carr (1998) later studied the effect of a priority-based scheduling algorithm in reducing the allocated deviations from the preferred airline arrival times. While these contributions demonstrated improvements in the capacity of the Traffic Management Advisor (TMA) system currently in place to improve fuel and throughput performance, their impact was limited since the system only operated out to a range of 200 nm. The aircraft sequencing problem attempts to deal with the congestion issue by improving terminal airspace throughput. The problem was first examined by Dear (1976), who studied the effect of constraining the movements of aircraft through constrained position shifting (CPS). More recent work Beasley et al (2000); Balakrishan and Chandan (2010) has resulted in efficient dynamic programming, integer programming, and heuristic approaches. Despite these advances, the focus of the aircraft sequencing problem has been oriented towards eliminating delay. Due to the heavy degree of congestion, optimal flight sequencing is very often insufficient to eliminate the need for the complex maneuvers described previously. In these cases it can be beneficial to transfer some of that delay to other phases of flight. An enhanced version of the TMA system called The Terminal Area Precision Scheduling and Spacing System (TAPSS) was then proposed (Swenson, et al., 2011). The technology has also been used cooperatively in Traffic Flow Programs (Grabbe, et al., 2012). Efforts with TMA have most recently been focused on an effort known as the Advanced Technology Demonstration (ATD-1) (Baxley, et al., 2013). The concept links TMA with Flight deck Interval Management Systems (FIMs) on-board the aircraft and Ground based Interval Management Systems (GIM-S) and Controller Managed Spacing (CMS) tools available to the controller. These systems interact to create greater situational awareness for both the controllers and pilots to better enable them to meet the desired spacing both en route and in the terminal. Carrier-centric approaches such as the Airline Based En route Sequencing and Spacing (ABESS) tool have also been proposed. The tool sends speed advisories to the Airline Operations Centers (AOCs) to allow crews to more actively manage their speeds en route (Moertl, 2011). To function effectively, these methods require reliable avionics algorithms that will enable flights to meet their arrival times. To that end, Tino (2013) proposed an algorithm that incorporated wind forecast into a multi-stage stochastic programing model to aid Flight Managements Systems (FMSs) in meeting the Required Times of Arrival (RTAs). The use of speed control has also been considered in the descent phase of flight to provide improved sequencing and spacing of flights along optimal profile descent maneuvers (Lowther, et al., 2008). While these approaches represent significant steps toward application of speed control, they do not account for the role uncertainty plays in perturbing flight assignments. Our approach attempts to mitigate the effect of such perturbations by accounting for the presence of uncertainty prior to the assignment of CTAs.

Building on many of the same concepts, practitioners within industry have developed speed control programs to enhance their operational performance.

Airservices Australia developed the ATM Long Range Optimal Flow Tool (ALOFT) to allow pilots to control speeds out to 1000 nm away from the airport. In so doing, they achieved an estimated fuel savings of nearly 1 million kg in 2008 Airservices Austrailia (Airservices Austrailia, 2008). Since then, they have also used additional metering fixes to better manage trajectory and arrival time uncertainty (McDonald & Bronsvoort, 2012). Delta Airlines achieved an estimated $8 million in fuel savings over a 20-month period using a dispatch monitored speed control program known as Attila (Leib, 2008). At Schiphol Airport in Amsterdam, a ground-based planning system that interfaced with aircraft through datalink was used to remove airborne holding in their nighttime operations (Nieuwenhuisen & de Gelder, 2012). Knorr (2011) identified substantial inefficiencies in the terminal phase of flight and characterized the benefit pool that could be achieved by "transferring" terminal delays to the en route phase of flight. Jones (2013) developed a bi-criteria integer programming model to facilitate delay transfer away from terminal airspace and demonstrated that a substantial proportion of the potential delay transfer benefit could be realized through this approach. In that study, the model objective attempted to explicitly account for en route flight fuel burn at various speeds and balance that with a need it did not directly address: the role conflicting flight arrival times can have in producing airborne holding. McClain (2013) examined a similar problem proposed a stochastic programming model that accounted for arrival time uncertainty due to wind and disturbances related to pop-up flights. Unlike our approach, the solution presented did not, however, address the role the departure delay distributions plays in contributing to airborne queuing near the terminal.

Speed control measures have also been considered at the pre-tactical level. Delgado and Prats (2012) showed that it was possible to absorb some of the delay assigned to flights within a Ground Delay Program (GDP) while en route and maintain the planned level of fuel consumption. The same research team also showed that by departing earlier but flying at a slower speed, a considerable portion of the imposed delay could be recovered in the event of an early GDP cancellation (Prats & Hansen, 2011; Delgado & Prats, 2012; Delgado & Prats, 2014). These studies proposed various methods for dealing with capacity uncertainty but did not address the role demand uncertainty plays in affecting arrival times. Delgado and Prats (2013) also considered the effect of wind forecast errors on the ability of flights to meet their assigned arrival times. The authors proposed adjusting calibrated flight speeds from their original assignments as a means of recourse when the actual winds differed from the forecast; however, they did not incorporate predictions of wind uncertainty into the planning process. While these studies introduced some important ideas toward improving the functionality of the National Airspace System, their intended use was oriented toward situations in which the airport capacity is significantly compromised. In many instances, an airport can operate closer to standard capacity yet the demand from flights can slightly outstrip airport capacity, leading to less severe but still significant delay. Speed control can also be useful in these situations; in Chapter 2 we attempt to address this system deficiency.

1.2.2 Air Traffic Flow Management Models

The first instance of the air traffic flow management model was proposed by Odoni (1987). It examined the single airport ground holding problem (GHP). Since then the scope of research on this type of problem has been expanded to account for a number of different facets of the problem. These facets are: adaptability, connectivity, control, rerouting, equity, speed control, and uncertainty. A taxonomy of models is shown in Table 1.1.

Table 1.1: Classification Metrics for ATFM models

Feature	Classification	
Adaptability	Static	Dynamic
Connectivity	Single Resource	Multiple Resource
Control	Central Decision Maker	Collaborative Decision Making
Rerouting	without Reroutes	with Reroutes
Equity	Equity not Considered	Equity Considered
Speed Control	without Speed Control	with Speed Control
Uncertainty	Deterministic	Stochastic

In addition to the ground holding problem, single-resource ATFM models have been developed to study airports in the Single Airport Ground Holding Problem (Terrab & Odoni, 1993; Richetta & Odoni, 1993). These models minimized the ground holding costs for flights over a single airport. Extensions to this work have been formulated by Hoffman and Ball (2000) to include banking constraints which require that groups of flights arrive within specified time windows.

Multiple resource models later evolved from the work on single airports models to consider flight ground holding (Vranas, et al., 1994). The model considered the connections between aircraft operating multiple flights. Modeling in this area quickly evolved to consider both ground and airport holding in a network setting. The Bertsimas and Stock-Patterson (1998) model used novel variable definitions to extend the work to the regional air traffic management system. This variable definition was useful in that it led to flight and network connectivity constraints that were in many cases facet defining. The model proved a significant advancement and allowed ground and airborne holding as well as speed control but due to computational limitations, it was not able to perform rerouting. The model was later revised by Bertsimas, Lulli and Odoni (BLO model) using a stronger formulation in that it was closer to the convex hull of feasible solutions (Bertsimas, et al., 2011). This update allowed rerouting to be incorporated into the model. Agustín et al., (2012a) developed a deterministic model which incorporated flight cancellations as well as speed control, rerouting, and airborne and ground holding. A companion paper Agustín et al., (2012b) also proposed a stochastic version of the model that accounts for capacity uncertainty as well as flight demand uncertainty in the form of flight cancelations from carriers. The model did not, however, consider flight arrival time uncertainty. A separate line of models attempt to optimize aircraft trajectories by incorporating equity, controller workload and probability of conflict (Sherali, et al., 2003; Sherali, et al., 2006).

The aforementioned models each attempt to manage resources from the point of view of a single decision maker. Another line of models incorporates CDM philosophy into their decision-making. Vossen and Ball have sought to improve the compression aspect of CDM by proposing a model to facilitate inter-airline slot trading (Vossen & Ball, 2006a; Vossen & Ball, 2006b) in a single airport setting. These models allow airlines to submit at-most/at-least offers meaning if an airline controls two flight f_1 and f_2 will delay flight f_1 at-most n slots in exchange for the ability to move up flight f_2 at-least m slots. The mechanism allows airlines to trade slots and improve the airlines' ability to optimize their cost functions. The model structure was later extended to a network setting by Gupta (2012). The APCDM model has also been extended to create opportunities to for slot trading between airlines (Sherali, et al., 2011).

The Vossen and Ball, BLO and APCDM models each consider equity in their approach. In Vossen and Ball (2006a), the authors show that their OPTIFLOW integer programming model, which uses cost coefficients that grow super-linearly with delay, has properties similar to RBS. The BLO model incorporates similar cost coefficients in their model to incorporate equity. The APCDM model includes a term in the objective function that penalizes/rewards airlines for assigning flights that deviate from some mean value of collaborative efficiency (Sherali, et al., 2003). Lulli and Odoni (2007) demonstrated that in some cases, when attempting to use RBS in a network setting, it may not be possible to achieve the most efficient solution. To address this issue, Churchill (2010) proposed a model designed to balance the need for equity with efficiency. Barnhart et al., (2012) proposed two other models designed to accomplish the same end.

There has also been significant development in the realm of stochastic ATM models. This work began with formulation of the static stochastic single airport ground holding problem by Richetta and Odoni (1993). Ball et al, (2003) proposed another model for the problem to determine the optimal planned airport arrival rate under uncertainty. Mukherjee and Hansen (2007) later formulated a dynamic stochastic integer program to solve the problem. Gupta (2012) used a robust optimization framework to address capacity uncertainty in the ATFM problem. Ball et al., (2010) proposed an algorithm known as Ration-by-Distance, wherein the authors show that the algorithm minimizes the expected delay if a GDP is cancelled before its planned end time. Glover and Ball (2010) formulated a two-stage multi-objective optimization model to address trade-offs between efficiency and equity. Another model designed to address capacity uncertainty in AFPs was presented in Ganji et al., (2009). A stronger formulation of the same problem was shown in (Ball, et al., 2011). Churchill and Lovell (2011) proposed a model to address capacity uncertainty in a network setting.

On the adaptability axis, the problem can be view in either a static sense in which the problem is solved once and any new information that materializes after the solution is obtained is ignored, or a dynamic case in which the problem is continually resolve to incorporate new information. While there has been some work in the dynamic area, there have been far fewer contributions than in the static case. The Mukherjee and Hansen model examines GDPs from a dynamic perspective; however the model does not leverage speed control as a mechanism for improvement. The APCDM model uses speed control to plan for conflict uncertainty. This model is largely intended for strategic planning; however, it can be used for tactical planning. It seeks to identify trajectories that minimize fuel and delay while accounting for controller workload, equity and safety constraints. While the model manages departure times and can be solved iteratively, it does not incorporate speed control for en route flights.

1.3 Background

1.3.1 Fuel Savings Assumptions

As discussed above, it is not unusual for a variety of flight maneuvers to be used in the terminal area of an airport in order to organize the flow of traffic into the airport into an efficient pattern. The most typically used maneuvers are vectoring in which a flight veers off course to a waypoint and the return to its initial planned trajectory and the race track shape patterns mentioned earlier, as well as long “downwind” approach paths (also called “tromboning” – see Figure 1.4). All of these techniques, which are used to delay the arrival time of a flight, represent path extensions; i.e. they add to the total distance flown by the aircraft. The extra distance was not the goal, however; rather the goal was to add extra time to the trajectory. Another way to accomplish the same thing is to reduce the aircraft speed as it approaches the airport. Our contention is that this action can be taken when the aircraft is at a higher altitude, which leads to a reduction in fuel consumption.

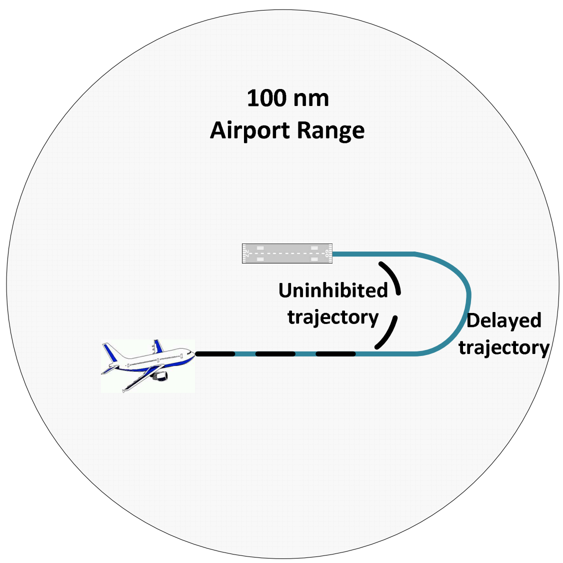

Figure 1.4 "Downwind" trajectory to absorb terminal area delay

Figure 1.5 illustrates notionally the relationship between the fuel efficiency (specific range) of an aircraft and its Mach number – the ratio of the speed of the aircraft to the speed of sound in air (Airbus, 2004). As the aircraft’s Mach number increases from zero, its fuel efficiency increases up to a point known as the maximum range, beyond which it begins to decline. The shape of this curve, importantly, is relatively flat in the vicinity of the optimum. This implies that one could fly at any speed within the flat part of the curve and use nearly the same amount of fuel for a given distance traveled. Shorter distances, therefore, imply less fuel burn.

Note also from Figure 1.5 that as altitude increases the specific range curves move markedly upward. Since the magnitude of the upward shift of the specific range is large relative to the increases along an individual curve at constant altitude, fuel efficiency at a high altitude is greater regardless of whether the Mach number changes significantly. This implies that if, as is typical, excess distance in the terminal airspace is taken at lower altitudes, then the fuel burn rate is higher than would be the case for a similar distance at a higher altitude. Thus, there are two effects at work that produce fuel cost savings when delay is transferred from the terminal area to the en route portion, though the reduction or elimination of path extension is more profound.

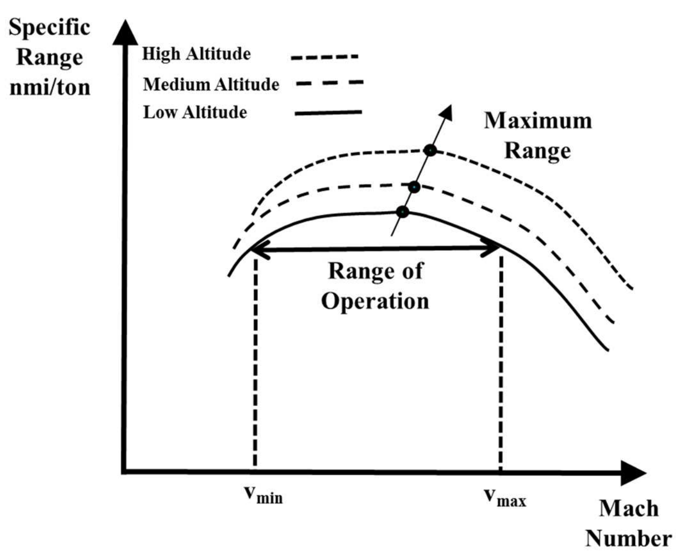

Figure 1.5 Notional variation in aircraft fuel efficiency with speed at various altitudes.

A significant feature of speed control assignment is that it may not be necessary to always impose delay on flights to achieve the benefits. Consider the following example depicted in Figure 1.6 and Figure 1.7. Here a group of flights in arriving is a set of consecutive slot times such that the rate of arrival exceeds the capacity of the airport. In the absence of speed control the flights would be forced to incur a delay of 9 minutes. By incorporating speed control into the set of control options we can assign each flight one slot earlier we and eliminate 6 minutes of delay.

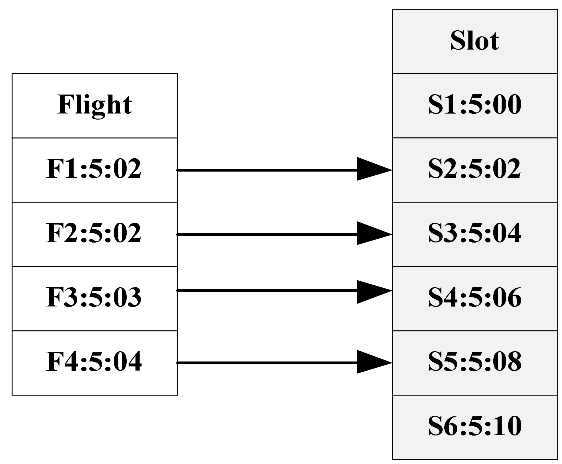

Figure 1.6 Potential Assignment in the Absence of Speed Control.

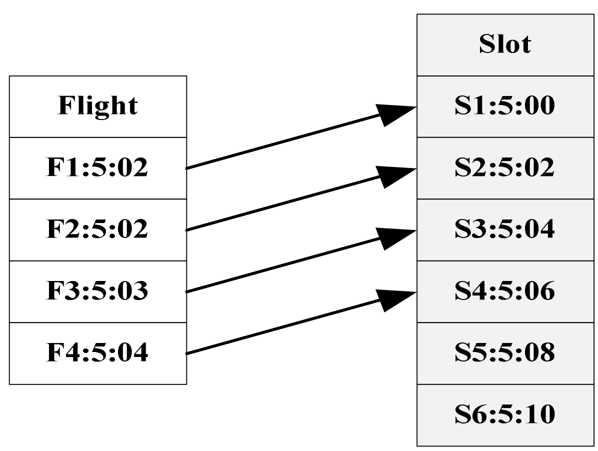

Figure 1.7 Potential Assignment with Speed Control.

1.3.2 Operational Concept

In the proposed operational scheme, CTAs are assigned to flights once they reach an approximate distance of 500 nm from the destination airport. Once a CTA has been assigned to a flight, it proceeds to the assigned metering fix 150 nm from the airport while exercising the appropriate speed control guidelines. When the flight reaches the metering fix, TMA would issue adjustments to controllers to effectively guide the flight on its assigned STAR trajectory. Under this concept, the system does not require close coordinate with TMA.

It is important to recognize that this is not a static problem. The changing environmental conditions necessitate that any assignment algorithm make use of revised information as it is presented. As flights travel en route, their estimated times of arrival (ETAs) are updated on an ongoing basis to account for factors such as changing winds, convective weather and rerouting. As flights get closer to their destinations, these ETAs become increasingly reliable. The ETAs provide a forecast of the degree of congestion and the resultant excess flight time and maneuvering that will occur in the terminal area. The assignment of CTAs effectively adjusts the ETAs to provide a more orderly flow of traffic into the terminal area, thereby injecting an increased level of predictability into the flow of traffic. In the longer term, the 500 nm horizon could be lengthened and also could vary by flight.

Under this approach, the Air Navigation Service Provider (ANSP) would update the list of flights that were available for scheduling every 15-30 minutes. At each period, the ANSP would set the number of "slots" at the metering fix based on the capacity of the airport and each metering fix. When the number of slots has been determined, an optimization model assigns a CTA to each flight once it reaches the 500 nm boundary.

These CTAs could be assigned using the various communications tools discussed in the following section. When the pilot receives this CTA he/she would enter this time into the Flight Management System (FMS) onboard the aircraft. The aircraft then calculates the preferred route and speeds en route and proceeds to the metering fix where it receives TMA-based controller instructions. It is important to note that the assignment process is iterative and dynamic. At each period, a new set of flights between 1-30 minutes away from the 500 nm boundary is evaluated by the assignment algorithm. Once the set of CTAs has been decided based on our model's logic, the flights receive a CTA only once they approach the 500 nm boundary. Note that there will generally be overlap between the set of flights considered from one iteration to the next as only the closest-in flights are given the computed CTAs. Thus, the CTAs computed for the further-out flights are temporary; these flights are included to provide an assignment procedure with a more global perspective of total flight demand.

1.3.3 System Description

For U.S. implementation, we anticipate that the Air Traffic Control Systems Command Center – ATCSCC – would have responsibility for determining the CTAs due to need for coordination across sectors. It is also the case that the data required to support these decisions are already readily available to the Command Center. The existing traffic flow management system (TFMS) integrates real time flight information such as estimated arrival times, scheduled arrival times, landing times, flight, aircraft positions and flight cancellations. The Command Center also has rich weather feeds and through consultation with airport Air Traffic Control Towers

(ACTs) and Terminal Radar Approach Control facilities (TRACONs), up-to-date information on airport and terminal area capacities.

In the longer term, CTAs would certainly be transferred to aircraft using datalink. However, this option will most likely not be possible in the shorter term. Thus, after examining the existing communications technology between pilots and command centers, we see two options for assigning CTAs. In the first, the Command Center passes CTA assignments to the Air Route Traffic Control Center (ARTCC), who informs the pilots of these assignment times via controllers / radio communication link. In the second approach, the Command Center communicates CTAs to appropriate airline operational control centers (AOCs). It is possible (at least in the longer term) that CTAs could be adjusted based on Command Center / AOC negotiation. Once a CTA was finalized, the AOC would send it to the appropriate aircraft over the Aircraft Communications Addressing and Reporting System (ACARS). Notionally this approach has the advantage of very naturally supporting the inclusion of (future) Collaborative Decision Making (CDM) features. The system may also integrate the TBFM and the extended metering system that has recently been adopted by the FAA. (Witzberger, et al., 2014)

The first approach offers a significant advantage from a compliance standpoint. Since assignments are issued directly by air traffic controllers, they will likely be taken quite seriously. Further, this approach would by necessity offer a degree of coordination between the CTA directives and other controller directives, e.g. those emanating from TMA. This approach could, however, impose an additional workload burden on some air traffic controllers and increase training needs at certain control centers. The second approach minimizes the burden on ATC staff by limiting their direct involvement in the assignment process. Although the ATC staff may issue resource capacity guidelines and updates to command centers to inform them in their decision-making, the assignments would be made jointly by carriers and the Command Center. This process allows the carrier to actively voice their priorities during the assignment process and potentially adjust their assignments through CDM mechanisms. The price of such accommodation, however, may be borne at the expense of operational effectiveness. If compliance is sufficiently low, it will likely prove quite challenging to realize a substantial portion of the potential benefit pool. Thus, it is critical that carriers actively enforce CTAs on their flights. Figure 1.8 and Figure 1.9 illustrate the flow of data between systems and stakeholders.

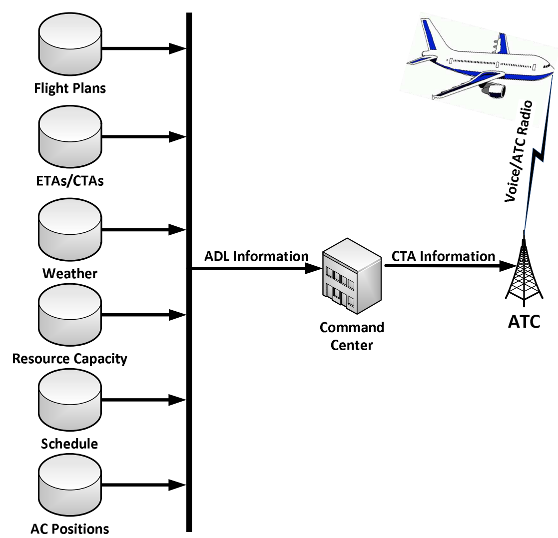

Figure 1.8 Information flow between databases, aircraft and command centers under a centralized approach

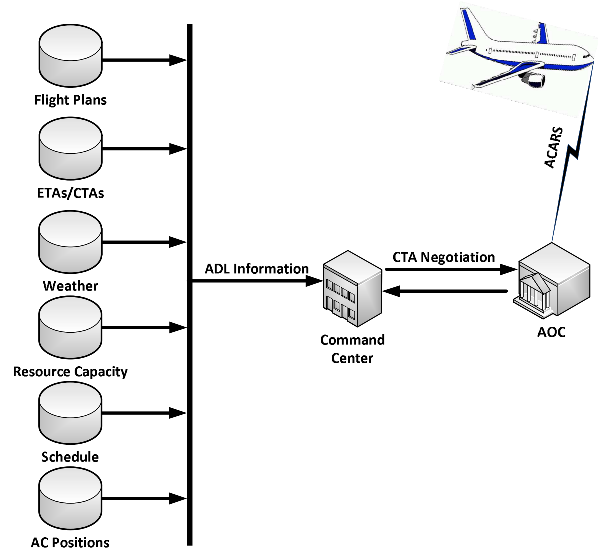

Figure 1.9 Information flow between databases, aircraft and command centers with collaboration from carriers.

1.4 Issues with current CDM Practices

1.4.1 Issues with collaboration

When a GDP is issued at an airport, air traffic managers at the ATCSCC decide the planned capacity and duration of the GDP based on the predicted conditions over the course of the day. They also determine the radius of exemption for the GDP. This exemption radius defines the set of flights that will receive ground delays. Thus there are three pools of flights to consider: flight inside the exemption radius receive ground delays based on their order in the schedule. Flights on the ground outside of the radius are exempted from the GDP and receive no delays. In addition, all flights already in the air and international non-Canadian flights, regardless of their origin, are exempted from the GDP.

Figure 1.10 illustrates an example RBS allocation where the two exempt flights identified on the left are both airborne at the time of allocation. After the RBS allocation, carriers may freely substitute flights based on their own priorities. They may also choose to cancel flights and make substitutions using the vacated slots. A notional example of this process is shown in Figure 1.11. Here, AA has chosen to cancel AA561 and move AA321 into its slot. AA alternatively could have chosen to swap the slots of the two flights. In either case, once the appropriate arrival changes were made, the arrival times (CTAs) would be converted to departure times (CTDs) and appropriate ground delays. Although DAL and UA both also have two slots, they are unable to make any changes since in each case, one of their two slots is occupied by an airborne flight whose arrival time cannot be adjusted. If DAL and UA could reassign the slots of their airborne flights, then each airline could improve the number of flights arriving less than 15 minutes after their scheduled arrival time. An example of this exchange is shown in Figure 1.12.

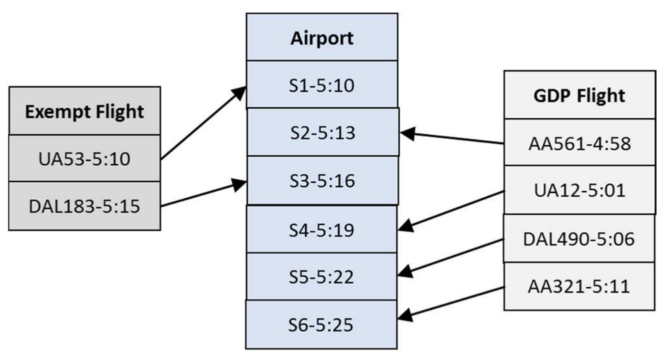

Figure 1.10 An example of flight allocation in Distance Based RBS. Exempt flights receive priority.

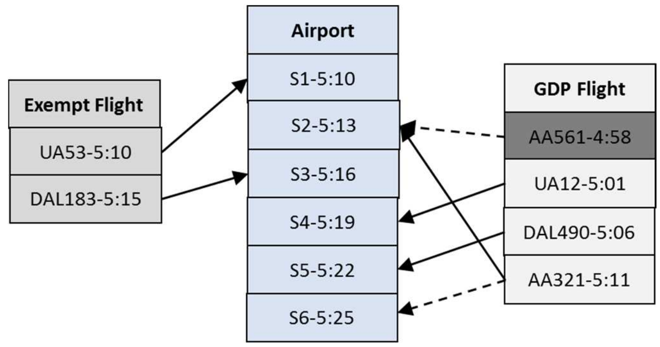

Figure 1.11 Cancellation and substitution procedure in the current CDM framework.

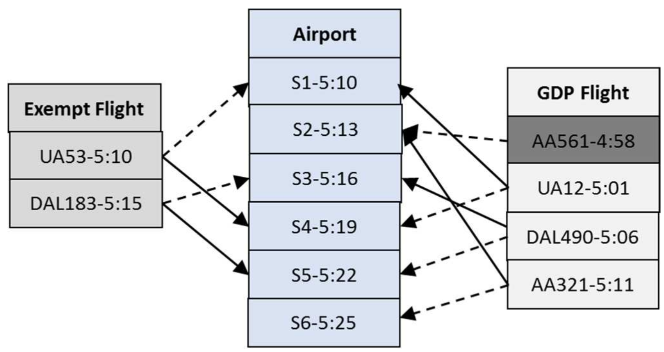

Figure 1.12 Cancellation and substitution process without exemptions. Delta and United can substitute and improve their on-time performance.

The advantages illustrated by these examples underlie the one component of the benefits that can be achieved by combining control by CTA with dynamic speed adjustments. Similar improvements to the performance of the compression algorithm can be achieved by allowing adjustments to airborne flights. We note a second source of benefits have the same origin as those investigated by Delgado and Prats (2012) and Delgado and Prats (2014), namely the ability of airborne flights to more quickly react to increases in arrival capacity resulting from weather changes

1.4.2 Issues with information and control

While RBS serves as a standard for equity within the system, it is not widely adopted once a flight leaves the ground. Indeed, controllers are allowed to use their own discretion to route flights between sectors. These judgment calls may often have no relation to the actual schedule. The algorithm, however, does not account for the uncertainty in meeting arrival times. Flights often deviate from their schedule for a host of reasons including convective weather, wind uncertainty and the availability of direct routing. These deviations can lead to scheduling conflicts when the flights reach the airports.

Consider the example depicted in Figure 1.13 and Figure 1.14. Here, three flights with scheduled times of arrival (STAs) within the immediate vicinity of flight F4all approach the airport at the same time. When this happens, three of the four flights will need to hold in the air until the airport can accommodate them. This additional holding leads to excess fuel burn on flights. Moreover, since air traffic controllers are not involved in managing TMIs, the actual order of precedence between the four flights could deviate from the assigned order. To make matters worse, since carriers will often cancel flights to reduce the delays on the other flights they operate the lack of predictability in slot availability increases the risk of such activity and lowers the incentives for the behavior CDM was designed to promote.

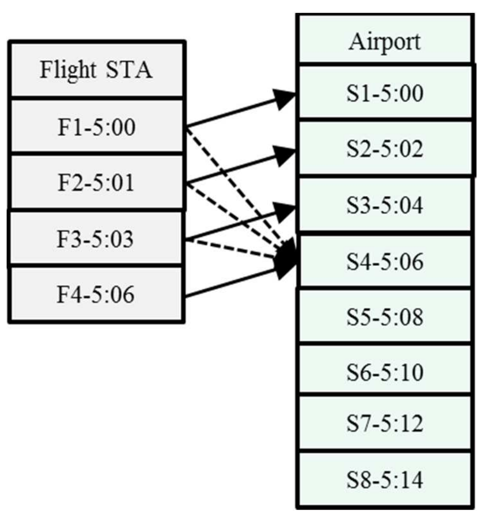

Figure 1.13 An illustration of scheduling conflict due to uncertainty in flight arrival times.

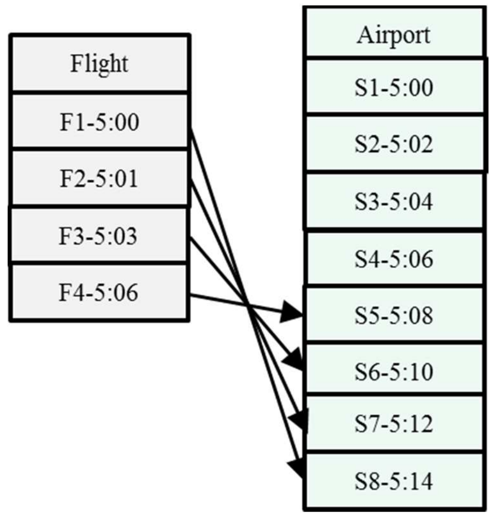

Figure 1.14 The effect of scheduling conflicts on flight arrival times

1.5 Research Contributions and Contents

This dissertation makes a number of contributions oriented towards advancing the use of en route speed control in air traffic management in the presence of uncertainty. This problem assumes an operational environment with less centralized control over system actors than what is assumed in the network-based Bertsimas (1998), (2011) and Agustin (2012a), (2012b) models, and assumes greater system uncertainty and more rapidly evolving conditions than in previous single-resource models which allow for greater decentralization. Thus, neither model is appropriate to solve this sort of problem. To bridge the gap, we provide a set of solutions to support both time-based metering and GDPs. Our contributions are as follows:

- To the best of our knowledge, this dissertation presents the first integer programming models to consider the role of both departure and arrival time uncertainty in coordinating flights en route. In this effort, we develop both a scenario-based and a value function-approximation based model, and we demonstrate their ability to transfer delay under dynamic environmental conditions. We show that value function approximation displays strong efficient solutions while achieving superior delay transfer, making it a particularly well-suited solution to the problem.

- The previous approach assumes a system in which the Air Navigation Service Provider (ANSP) assumes control of airborne flights but no ability to control flights prior to take-off. To add versatility, we introduce three new models which allow the ANSP to assign ground delay to flights prior to take-off. The models are stochastic and account for the uncertainty in departure delay. The first model is a two-stage scenario-based integer program that utilizes the “by-variables” originally introduced in (Bertsimas & Patterson, 1998) whose solutions to the LP relaxation were shown to be facet defining for the ATFM problem. The second model adapts to the functional approximation model presented in Chapter 2 into a system with stronger control over ground-based flights. We also explore an alternative functional approximation model that incorporates the by-variables and present some valid inequalities to improve the computational performance. The models demonstrate strong delay transfer with relatively little imposition of ground delay on short-haul flights.
- We propose the concept of applying speed control to reduce the GDP exemption bias. To support our claim, we modify the current models and conditions to reflect the proposed changes within our GDP procedure.

- The three previous approaches examine the problem of en route speed control from the perspective of the FAA. In a more decentralized version of the NAS, carriers are likely to play a more active role in managing their flights with speed control. To that end we propose a new GDP architecture that controls flights through CTAs as opposed to CTDs and gives carriers an increased level of responsibility over the flights they operate. To aid the carriers in their new planning responsibilities, we introduce a stochastic model to support airline decision-making during GDPs. The model allows carriers to hedge between the possibility of delay and the likelihood of early weather clearance. We show that by moving to this new architecture carriers could realize significant cost savings.

Chapter 2 presents three optimization models for en route speed control that can be used in concert to address the inherent uncertainty in flight arrival times. In Chapter 3, we add the ability to control ground based flights and propose a set of alternative models to the ones discussed in Chapter 2. In Chapter 4, we present two approaches designed to curb the exemption bias in GDPs using en route speed control. We later remove the exemption radius entirely and provide carriers with capability of using speed control to control arrival times and propose a stochastic model to support airline decision-making during GDPs. In Chapter 5, we present our conclusions and ideas for future work.

2 Transferring Delay through Control of Airborne Flights

In a typical aviation environment today, the precise landing times of en route aircraft are not set until each aircraft approaches the airspace adjacent to the destination airport. In times of congestion it is not unusual for air traffic controllers to subject arriving aircraft to various maneuvers so as to create an orderly flow of aircraft onto an arrival runway. Typical maneuvers might include flying in zig-zag patterns, flying in circular holding patterns, as well as others. These maneuvers serve to delay the arrival time of the flight. On the other hand, if the arrival time was established much earlier, then such delay could be added by simply having the flight fly slower while still at a higher altitude, which would incur much less fuel burn than the described maneuvers.

In this section we propose three integer programming models to assign delay to aircraft approaching a single airport, well in advance of each aircraft's entry into the terminal airspace. The baseline model is deterministic and seeks to maximize the available throughput at the runway over a rolling-horizon. The latter two models are stochastic and account for uncertainty regarding the status and controllability of certain flights. The first stochastic model is scenario-based, while the second relies on a functional approximation of uncertainty. The results of computational experiments show that these stochastic model approaches can transfer a considerable portion of the delay that would otherwise occur in the terminal area to the en route phase of flight and also that the stochastic models are noticeably more effective. The model relying on functional approximation shows particular promise due to its efficient run time. The delay transfer yielded by each model resulted in significant predicted fuel savings. The functional approximation model performed particularly well under declining operational conditions, demonstrating itself to be a promising means of achieving delay transfer.

2.1 Operational Approach

The goal of our approach is to adjust the speed of a flight during the en route portion of the flight so that when it arrives in the terminal area it will be able to land at the airport with little or no trajectory adjustment. This is accomplished by issuing to each approaching flight a controlled time of arrival (CTA). CTAs would be assigned at a notional boundary well in advance of the destination airport. This boundary imposes a limit on both the number of flights that can be controlled and the amount of delay that can be transferred through en route speed control. When the radius is too large, we cannot control a sufficient number of flights to make a strong impact. When the radius is too small, it is not possible to fly at the appropriate speed long enough to transfer much delay. With these factors in mind, we selected a boundary 500 nm from the destination airport. The CTA represents the time at which the aircraft should pass a metering fix (a defined point in the airspace) approximately 150 nm from the airport. When the flight reaches the metering fix, the controllers, using advice from TMA, would take over the final spatial and temporal control of the flight. Under this concept the system does not require close coordination with TMA.

It is important to recognize that this is not a static problem. The changing environmental conditions necessitate that any assignment algorithm incorporate new information as it is presented. Under this approach, the air navigation service provider (ANSP) would update the list of flights that were available for scheduling every 15-30 minutes. At each period the ANSP would set the number of "slots" at the metering fix based on the capacity of the airport and the capacity at each metering fix. When the number of slots has been determined, an optimization model assigns a CTA to each flight once it reaches the 500 nm boundary. When the pilot receives this CTA, he/she would enter this time into the Flight Management System (FMS) on board the aircraft. The aircraft could then calculate the preferred route and speeds and proceed to the metering fix, where it would then receive TMA-based controller instructions.

In our approach, flights can be grouped into two classes: long-haul flights originating at airports greater than 500 nm from the destination, and short-haul flights originating at closer distances. Long-haul flights are managed by assigning arrival times at 500 nm. Short-haul flights cannot be managed by the process until they reach a distance of 150 nm, at which time they begin to follow TMA-initiated instructions. At this radius of 500 nm, short-haul flights can compose a substantial portion of the flight pool, and at many airports make up the majority of the flights. As such, the uncertainty associated with the arrival times of these unmanaged flights plays a considerable role in determining the ability of managed flights to make their assigned arrival times. For example, if an unmanaged flight arrives a few minutes later than its ETA, its arrival time can overlap with that of a managed long-haul flight. When this occurs and the airport does not have the capacity to accommodate both flights, air traffic controllers are forced to hold one of the flights until it can be accommodated. An example is shown in Figure 2.1. Our aim in this dissertation, and the key factor distinguishing this work from other related published work, is to show that by accounting for the presence of demand uncertainty, it is possible to issue CTAs that are more effective in limiting the excess delay taken in the terminal area.

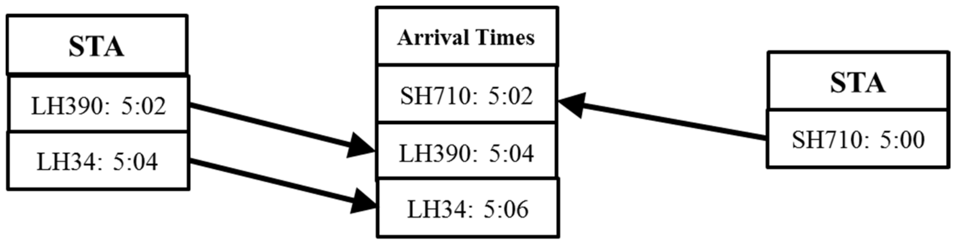

Figure 2.1 Short-haul flight (SH710) arrives 2 minutes late, delaying the arrival of two long-haul flights (LH390,LH34)

The approach taken herein is generally applicable to a wide array of demand conditions. On fair weather days when no Traffic Management Initiatives (TMIs) are in effect at the subject arrival airport, flight arrival times are affected by a wide array of factors beyond the control or knowledge of the controllers. During Ground Delay Programs and Airspace Flow Programs, while the initial plan is constructed using evenly spaced arrival time slots as a notional goal, the controls are executed at the departure stage of the flight, beyond which time many key flight parameters can still change. Thus, the arriving traffic stream, even under a TMI, can still be quite unpredictable and could benefit from final coordination. There is a limit to the magnitude of the incoming traffic flow; any serious imbalance between demand and capacity would be mitigated by the FAA instituting a TMI in response.

A natural extension of this work might be to also control the short-haul flights. In fact, this could potentially be done, although the characteristics of the control would be different; e.g. it could include delaying the flight's departure time. Our current goal, however, is to operate with very limited changes to existing air traffic management procedures so that the only new control required is the issuance of CTAs to flights when they are 500 nm from their destination airport. Accordingly we decided to limit the scope of intervention in this study to the flights originating at distances beyond 500 nm.

The stochastic models described in the next section explicitly model the uncertainty associated with short-haul flights. There are other sources of uncertainty related to the timing of the arrival of flights. First, there will be some variability in the speed and arrival time of the flights issued CTAs. This variability could be due to a number of factors, such as the inaccuracy in the estimate of wind characteristics over the course of the trajectory. Another, perhaps more significant source, is the degree of compliance with the issued CTAs and also the degree to which it will be feasible for the pilot to employ the CTAs calculated by the model. Depending on the manner in which the CTAs are conveyed to the pilot, he/she might not be obligated to comply with the CTA and other priorities might be given preference over adhering to the CTA provided. It is also the case that the technology on the aircraft might make it difficult to communicate the CTA and/or for the pilot to accurately make use of it. Of course, there could also simply be delays in the CTA communication and implementation. These sources of uncertainty are not explicitly incorporated into the models presented in the next section; however, certain experiments and model changes are used to study their impact in sections 2.3, 2.4 and 2.5.

2.2 Methodology

In this section we describe the structure of the three models introduced in this paper to assign arrival times to flights. All models assume a multi-resource framework in which the assignment times are issued at metering fixes 150 nm away from the airport and are compatible with available runway arrival times. The models iteratively re-solve the problem, in rolling-horizon fashion, to accommodate the changing conditions within the airspace. Each model aims to transfer delay away from the terminal; however, our first model assumes conditions are deterministic while the latter two incorporate different versions of a stochastic framework.

2.2.1 Basic Model Structure

As discussed in the previous section, the ultimate control variable of the system, a speed adjustment, is determined implicitly by assigning each flight a slot at a fix. As Figure 2.2 illustrates, however, the model must specify both a flight-to-fix assignment and a flight-to-runway assignment. Specifically, each fix will have a capacity (maximum flow rate) and similarly each runway will have a certain capacity (maximum arrival rate). These capacities are converted into slots; e.g., if a runway capacity was 45 arrivals every hour then 45 slots would be created in each hour, equally spaced. In general multiple fixes can feed multiple runways. Thus, all of our models include assignment variables that assign flights to both a fix and a runway. Of course, there are multiple ways to model these assignments within an integer program: one could employ two different sets of flight-to-slot variables with constraints ensuring compatibility between the fix and runway assignments, or one could use a single variable to assign the flight to a slot at both a fix and a runway. We experimented with both approaches but chose the latter because it produced superior computational performance. This is not surprising since the use of such "composite" variables is equivalent to moving from an "arc-based" formulation of a routing problem to a "path-based" or set-partitioning approach. Such formulations are known to be stronger (see, for example Chapter 11 of Wolsey, 1998). The usual disadvantage of such a transformation is that the formulations become very large; however, since we are effectively dealing with paths of length two (500 nm boundary-to-fix-to-runway), the number of variables in the composite formulation is quite manageable.

In our formulation, we use a set of parallel slot lists, each slot corresponding to a single fix-runway pair. As mentioned above, each slot can be occupied by at most one flight. Therefore, by assigning flights to these slots, we automatically ensure that the capacity constraints are enforced. Furthermore, as will be seen later, each runway queue is accounted for separately, and the objective function is to minimize the total queuing delay. This results in an assignment that is as closely balanced across fixes as possible.

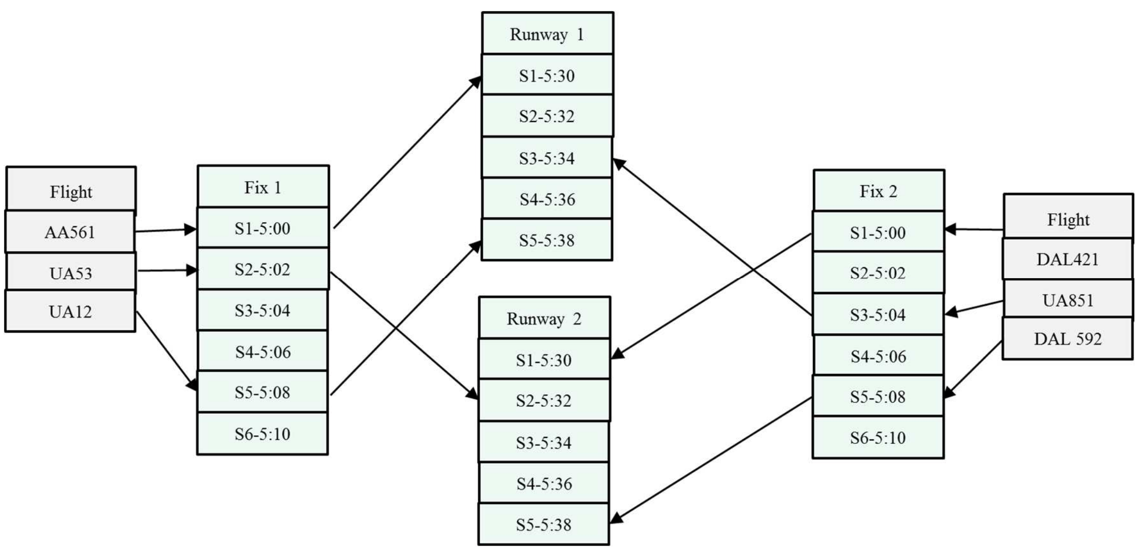

Figure 2.2 A notional representation of our model assignment structure.

2.2.2 Model Uncertainty

There are three principle dimensions along which the presence of demand uncertainty can affect traffic: (1)Flight cancellation uncertainty which related to whether a flight operates or not can affect the overall demand for an airport and influence the need to limit the capacity at the affected airport to curb imbalances. (2)Trajectory uncertainty relating to the ability flights to adhere to their planned/predicted route can alter arrival times both at the airport and metering fix, as well as introduce additional conflicts between aircraft. (3)Finally timing uncertainty regarding when flights depart from their origin airport and how long a flight takes to travel between waypoints and airports can significantly influence the level of airborne holding and produce additional conflict in airspace. While all forms of uncertainty can prove significant in controlling flights in the presence of convective weather, we shall limit the scope of the analysis to focus on timing uncertainty.

There are a number of ways to approach the uncertainty associated with the arrival time of the short-haul flights. Perhaps the most obvious deterministic model would involve using mean flight times to calculate fix arrival times for all short-haul flights. These flights would be pre-assigned to appropriate slots. The remaining slots would then be available for assignment to the long-haul flights. An alternative deterministic approach would be to ignore the short-haul flights in making the long-haul assignments so that the short-haul arrivals would simply be a stochastic event dealt with after the fact. While the first approach might seem more appealing, the second approach provided superior performance (in preliminary experiments not reported here) so it will be used as our baseline deterministic model.

Our baseline stochastic model is a scenario-based, stochastic, integer programming model that employs samples from a representative distribution of short-haul flight times. Of course, the accuracy of this model depends on the number of scenarios generated, with larger numbers of scenarios leading to increased model run times. For some static problems this increased solution time may not be an issue, as models do not often need to reach a solution over a short time horizon. In more dynamic cases the solution time becomes increasingly critical. In our application we need to achieve solutions and issue CTAs over a 15 minute time horizon. Some amount of buffer time is necessary to execute the instructions, after which we must move on to the next iteration of the problem. This time criticality raises the question of whether a scenario-based approach is most appropriate to our application. While we want to achieve a high quality solution, it may not be possible to do so by incorporating a large number of scenarios into the model.

To deal with this issue we propose a second approximate stochastic model. Specifically, the two stochastic models both employ a decision vector, *y*, which assigns long-haul flights to slots. Both models take into account the anticipated arrival times of short-haul flights captured by the variable vector *n*, which gives the number of short-haul flights whose planned trajectory and speed would result in arrival to the fix at each time slot. We denote by $f(y,n)$ the excess delay taken in the terminal area for a specific *y* and *n*. The first model is based on a set of short-haul flight arrival time scenarios, with each scenario *s*, characterized by a vector n^s. This model minimizes the expected excess terminal area delay, $E\left[f\left(y,n\right)\right]$, where the expectation is taken over the sampled scenario distribution. For the second approximate model, we compute a priori the expected value of *n*, $E[n]$, and then minimize $f\left(y,E[n]\right)$. We note that this model can be viewed as a functional approximation of the scenario-based model as

$f(\ )$ is not linear and the vector $E[n]$ can be non-integer. Thus, while the decision vector y is integer and explicitly assigns long-haul flights to slots, there is not an explicit assumption regarding when each short-haul flight arrives (as there would be in a more standard deterministic approximation). Conceptually, the first model is more accurate than the second since it explicitly minimizes the expected value, while the second, by moving $E[n]$ inside $f(\ )$, employs an approximate objective function. On the other hand, the accuracy of any scenario-based model depends on the degree to which the scenarios generated accurately represent the true distribution. Of course, to get more accuracy, more scenarios must be generated (leading to larger models). We explore these tradeoffs in our computational experiments and in fact show that the second model can be very effective.

The objective functions in the stochastic models described in the previous paragraphs deal with characterizing expected value. In some contexts the worse-case scenarios can cause significant impact to the system under study to the point at which it becomes more helpful to consider objective functions that use risk-based measures such as Value-at-Risk and Conditional Value-at-Risk (Gaivoronski & Pflug, 2005; Rockafellar & Uryasev, 2000). In air traffic management there are a number of measures put in place to prevent such incidents from occurring. It is also not our intent to develop a set of models to aid stakeholders under such catastrophic conditions. For these reasons we shall refrain from presenting any risk-oriented models and limit our discussion to the aforementioned models oriented toward expected value.

2.2.3 The Deterministic Model

Our deterministic integer programming model employs an objective function that minimizes total system delay. Since it assumes no variability in flight times and it ignores short-haul flights, it assigns each long-haul flight to a unique slot and thus is able to transfer all delay from the terminal area to the en route portion of the flight (each flight adjusts its speed so that it arrives exactly at the time of its assigned slot). This model considers flight assignments over a rolling two-period horizon by discounting the second period to a marginally lower level to account for a lower degree of confidence in more distant events. We know the information about the first period with more certainty, and we only make decisions about the first period. To the extent that decision variables are used for the second period, their role is to facilitate the assignment of aircraft whose ETAs were in the second period into first-period slots. The notation used to describe the data in the two periods is identical, but functionally the two periods play quite different roles.

In order to limit the number of constraints in our model, certain restrictions were imposed on some of the sets. Since it may be impractical for aircraft to periodically change their designated approaching corner posts throughout the course of flights, we restricted the assignment of each flight to its planned fix at 500 nm from the airport. In order to ensure that flights did not operate at unsafe speeds we also restricted the range of slots over which each flight could be assigned to times corresponding to either Mach 0.72-0.85 or the performance of the aircraft, whichever criterion was more restrictive. We define our variables and parameters as follows:

<u>Parameters</u>

$F \equiv$ The set of all flights

S_{if} ≡ The set of all slots available to flight f at fix i

Y_{rf} ≡ The set of all slots available to flight f at runway r

Ω_f ≡ The set of all fixes available to flight f

T ≡ The set of all periods

R ≡ The set of all runways

t_{sr}^{j} ≡ the time corresponding to slot s at runway r during period j

$e_{fsr}^{\ j}$ ≡ the earliest possible time flight f can be assigned to slot s runway r during period j

c_{fsr}^{j} ≡ the cost of assigning flight f to slot s at runway r during period j

β ≡ the discount factor for the second period of the rolling-horizon, where $\beta \leq 1$

<u>Variables</u>

$$x_{fks}^{irj} = \begin{cases} 1 \text{ if flight } f \text{ is assigned to slot } k \text{ at fix } i \text{ and slot } s \text{ at runway } r \text{ during period } j \\ 0 \text{ otherwise} \end{cases}$$

The deterministic model can then be stated as follows:

$$\min \quad \sum_{\substack{f\in F, k\in S_{if}, \\ i\in\Omega_f, s\in Y_{rf}, \\ r\in R}} \left(c_{fsr}^{1} x_{fks}^{ir1} + \beta c_{fsr}^{2} x_{fks}^{ir2} \right) \tag{2.1}$$

$$\sum_{\substack{k\in S_{if}, i\in\Omega_f \\ s\in Y_{rf}, r\in R, j\in T}} x_{fks}^{irj} = 1 \quad \forall f \in F \tag{2.2}$$

$$\sum_{\substack{f\in F, \in Y_{rf}, \\ r\in R}} x_{fks}^{irj} \leq 1 \quad \forall k \in S_{if}, i \in \Omega_f, j \in T \tag{2.3}$$

$$\sum_{\substack{f\in F, k\in S_{if}, \\ i\in\Omega_f}} x_{fks}^{irj} \leq 1 \quad \forall s \in Y_{rf}, r \in R, j \in T \tag{2.4}$$

$$x_{fks}^{irj} \in \{0,1\}, \forall f \in F, k \in S_{if}, i \in \Omega_f, s \in Y_{rf}, r \in R, j \in T \tag{2.5}$$

Equation (2.2) states that every flight is assigned to one slot over the two time periods. Equation (2.3) states that each slot at each fix can be assigned to at most one flight. Equation (2.4) states that each slot at each runway can be assigned to at most one flight. Equation (2.5) states that the decision variables are binary.

Equation (2.1) seeks to minimize system delay over two periods and discounts the second period. The throughput coefficients will vary based on the amount of time between their corresponding slots and earliest possible arrival times. A more explicit expression of the cost coefficients is shown in equation (2.6).

$$c^j_{fsr} = t^j_{sr} - e^j_{fsr} \qquad (2.6)$$

There are a number of indices on our parameters and variables which could imply that the size of our problem is relatively large. We are able to reduce the size substantially through the manner in which we populate our sets. The aircraft performance and fuel cost curves limit the range of speeds at which the flight can travel to a small subset within the vicinity of the nominal aircraft speed. Thus when a flight reaches the 500 nm boundary, the range of reachable arrival times and correspondingly slots at the fixes and runways is fairly small. Using the real time ETAs collected when the flight reaches the 500 nm boundary and assuming that the flight would meet its ETA by traveling at the nominal aircraft speed, we can project the earliest possible arrival time by measuring the deviation between the travel times at the nominal and fastest aircraft speeds and subtracting that deviation from the ETA.

Theoretically, one could imagine pushing this model until it was infeasible, by introducing more demand than the available capacity could accommodate, over an extended period of time. The same can be said of the following two stochastic models. In reality, if this or any such model were implemented in the real system, such events would not come to pass, because controllers are on guard for such demand-capacity imbalances. If they threaten to appear, then Ground Delay Programs, Ground Stops, and other Traffic Management Initiatives are available to rectify the imbalance. Our models would be able to function with the reduced demand any such program would permit, providing speed control adjustments to coordinate arriving traffic.

2.2.4 The Scenario-based Model

While the deterministic model presented in the previous section treated the problem as one of assigning flights to slots, we could also view this problem in the context of lot-sizing. One can imagine an inventory holding problem in which the user is trying to determine a production plan over a set of periods. In this framework we can produce inventory (flights) in a specific period and store inventory over a period (airborne holding). In each period the resource (runway slot) has a demand of one flight. Since the flight schedules are defined prior to our involvement we can view the production costs as negligible. Thus we are challenged with the task of determining a production schedule that will minimize the airborne holding costs over all periods subject to our stock conservation constraints in each period. If we could completely control the number of flights assigned to each period we could solve this inventory holding problem using a deterministic model and determine a CTA schedule that would minimize airborne holding. Since we have no control over the short-haul flights, however, the number of flights in each period is a stochastic quantity. As such we decided to turn to stochastic programming to handle the problem.

We developed a scenario-based model designed to explicitly take into account the impact of the (non-controlled) short-haul flights. Since the arrival times of the short-haul flights are uncertain, the model must explicitly consider terminal area delay. In fact, the objective function of the model is to minimize expected terminal area delay and thereby maximize the expected amount of delay transferred to the en route portion of the flights. The model samples from a set of scenarios that represent the arrival times of the short-haul flights. The number of short-haul flights that arrive in each slot is computed for each scenario. The model includes constraints that calculate the number of flights queued under each scenario – the sum of these queue lengths over time represents the total delay the model would have assigned in the terminal area under that particular scenario. The objective function represents the expected value of this measure over all scenarios, which is the total expected terminal area delay. Note that the scenario-based model no longer employs two periods, since the randomness associated with short-haul flight arrival times has been encapsulated in the scenarios. To the extent that any remaining equation references previous equations, any dependence on the subscript representing time period should be suppressed. The resulting stochastic integer program can be seen below:

<u>Additional Parameters</u>

$p_q \equiv$ The probability of scenario q

$Y_r \equiv$ The set of all slots on runway r

$n_q^{sr} \equiv$The number of short-haul flights arriving in slot s at runway r in period j under scenario q

Note that while n_q^{sr} is a non-negative integer, it can take on values larger than one. Since we cannot control them, more than one short-haul flight can arrive in time to occupy the same slot. Furthermore, that particular slot may be assigned to a long-haul flight, in which case none of the short-haul flights will be able to use it. Thus, one can expect a queue to form, and we would like to make the delay impact of this queue as small as possible.

Additional Variables

$$y_{sr} \equiv \begin{cases} 1 \text{ if slot } s \text{ at runway } r \text{ is assigned a long haul flight} \\ 0 \text{ otherwise} \end{cases}$$

$W_q^{sr} \equiv$ The number of flights queued for slot s on runway r in scenario q

$$\min \sum_{q \in Q} p_q \sum_{s \in Y_r, r \in R,} W_q^{sr} \quad (2.7)$$

$$\text{s.t.} \quad (2.2-2.5)$$

$$x_{fks}^{ir} \le y_{sr} \quad \forall f \in F, k \in S_{if}, i \in \Omega_f, s \in Y_{rf}, r \in R \qquad (2.8)$$

$$W_q^{sr} - W_q^{s-1r} - y_{sr} + 1 \ge n_q^{sr} \quad \forall s \in Y_r, r \in R, q \in Q \quad (2.9a)$$

$$W_q^{0r} \equiv \hat{W}_q^{r} \qquad \forall r \in R, q \in Q \qquad (2.9b)$$

$$y_{sr} \in \{0,1\} \qquad \forall s \in Y_r, r \in R \qquad (2.10)$$

$$W_q^{sr} \ge 0 \quad \forall s \in Y_r, r \in R, q \in Q \qquad (2.11)$$

Equation (2.8) ensures that if a flight is assigned to a slot at a runway, then the occupancy variable y for that runway is set accordingly. Equations (2.9) define the overflow of flights into each slot, which ultimately defines the level of airborne queueing delay in each runway slot. An intuitive way to understand equation (2.9a) is to reorganize it to represent the queue dynamics:

$$W_q^{sr} - W_q^{s-1,r} \ge n_q^{sr} - (1 - y_{sr})$$

This version of the equation highlights how the queue can change size from one slot to the next. In each new slot, the number of new short-haul flights arriving contributes to the queue length. If that slot had been reserved for a long-haul flight, then $y_{sr} = 1$, so nothing subtracts from the queue length. If, on the other hand, $y_{sr} = 0$, then that slot was not reserved for a long-haul flight, so it can be used for one of the queued flights, and the queue length is decremented by one. Equation (2.9b) assures that the starting queue length for any iteration of the problem is equal to the ending queue length from the previous horizon, with $\hat{W}_q^r \equiv 0$ at the beginning of the day. Equation (2.10) says that our indicator variable is binary. Equation (2.11) requires that the queuing delay in each slot be non-negative. Our objective function seeks to minimize the queueing delay over all scenarios.

As will be seen in a later section, we were able to test different versions of the scenario-based model, with the number of scenarios ranging from a few hundred up to nearly 2000. Because each scenario coded in the IP represents a single sample path from the set of underlying distributions, the probabilities assigned to the scenarios are all uniform. One could imagine different processes for generating scenarios, however. For example, it would be possible to generate a large number of sample paths but then cluster those into "representative" scenarios, with their resulting probabilities. This approach is typically motivated by run time considerations; a stochastic description with just a handful of scenarios is better than none. In our case, however, we were able to embed hundreds or thousands of scenarios into the stochastic IP and maintain reasonable solution times. Our notation for the objective function (2.7) includes the probabilities associated with the scenarios, although, as was already mentioned, we used uniform probabilities. The notation is useful because other users of the model might prefer a different method for generating scenarios, in which case differing probabilities might be appropriate.

2.2.5 The Functional Approximation Model

The Scenario model explicitly accounts for the uncertainty of the unmanaged flights by modeling its behavior through monte carlo sampling. If the simulation presents an accurate depiction of the arrival process for the short-haul flights, it can serve as an effective means of incorporating demand uncertainty into the model. It could, however, require a substantial number of scenarios to model the processes accurately. Thus, it is worth exploring the possibility that other representations of the uncertainty may serve as better proxies. Specifically, we propose a functional approximation of uncertainty that uses the same distributions used in the scenario model to compute the probability that a given flight will be in each slot and sums the probabilities to compute the expected number of unmanaged flights in each slot. This value can then be used to calculate the queuing delay. The proposed model is shown below:

Additional Parameters

$n^{sr} \equiv$ The expected number of short-haul flights arriving in slot s at runway r

Additional Variables

$$y_{sr} \equiv \begin{cases} 1 \text{ if slot } s \text{ at runway } r \text{ is assigned a flight} \\ 0 \text{ otherwise} \end{cases}$$

$W^{sr} \equiv$ The expected number of flights in slot s on runway r

$$\min \sum_{s \in Y_r, r \in R} W^{sr} \qquad (2.12)$$

$$\text{s.t.} (2.2 - 2.5), (2.8), (2.10)$$

$$W^{sr} - W^{s-1r} - y_{sr} + 1 \geq n^{sr} \quad \forall s \in Y_r, \forall r \in R \quad (13a)$$

$$W^{0r} \equiv \hat{W}^{r} \qquad \forall r \in R \qquad (13b)$$

$$W^{sr} \geq 0 \ \ \forall s \in Y_r, \forall r \in R \qquad (14)$$

Equations (2.13) track the queueing delay in each slot, with the correct accounting across adjacent time horizons. Equation (2.14) ensures that this queuing delay is non-negative. Our objective seeks to minimize the aggregated expected queueing delay.

One might ask why this particular approximation for the expected delay was used. The function has two significant properties: when the number of runways is limited to one, the expected delay represented in the objective function as well as equation (2.13) serves as an exact transformation of the expected delay when all slots are occupied. When some slots are unoccupied, the functional approximation serves as a lower bound on the expected delay at optimality. A justification for these properties is provided in Appendix A. In instances where multiple runways are considered, the model only provides an approximation of the true expected delay. This discrepancy is due to the fact that the approximation evenly distributes the expected number of short-haul flights arriving during a given time interval (slot) over each runway. In the operational environment these short-haul flights are managed and controllers assign these short-haul flights to runways along with the long-haul flights. Thus the expected delay is a function of the decisions made on both the short-haul and long-haul flights. This treatment is captured in the scenario models at the expense of additional computational time; however, it only approximates the expected number of flights arriving in the given time interval.

2.3 *Managing Flights with Certainty*

A computational experiment was constructed to compare the performance of our models. A scenario set was constructed using historical flight data to study the effect of speed control measures at a single airport in the presence of demand uncertainty. As discussed at the end of section 2.1, there are other sources of uncertainty not explicitly incorporated into the models discussed in section 2.2. We investigate the impact of these sources of uncertainty as part of our experiments. In this section, we outline our procedure for generating the uncertainty, describe the scenarios and the associated assumptions, we present our experimental results, and we provide some analysis to compare the tested models.

2.3.1 Experimental Description

The basis of our experiment is a dataset collected from Atlanta Hartsfield-Jackson Airport (ATL) on May 1, 2011. This day could be described as a "fair weather day" since no Traffic Management Initiatives were deployed at the airport. The dataset was obtained by merging data from two sources: a Traffic Flow Management System (TFMS) file and an ASDX file (surface surveillance data). The key fields included: flight number, collection time stamp, expected time of arrival (ETA), origin airport and actual time of departure, current aircraft position, aircraft type, runway arrival time, and cornerpost fix. We assumed an airport acceptance rate of 100 flights per hour. This assumption is consistent with ATL operating practice under the weather conditions for the time in question (full use of 2 runways and partial use of a third). The experiment was run over a 4-hour period from 1:00-5:00 EST.

ATL has 4 cornerposts at the northeast, northwest, southeast and southwest corners of the airport. Arriving flights commonly fly through one of these cornerposts and are sent to one of 3 runways: 2 runways are used full time and another runway is partially used. The runway capacity is bound by the wake vortex separation requirement between classes of aircraft. Based on the general fleet mix present in the data we found that we could assign uniform slot sizes that could be adjusted later to achieve tighter spacing.

We developed a simulation intended to model the basic effects of TMA and more generally the manner in which the CTAs produced by our model would impact operations in practice. An illustration of the simulation framework is shown in Figure 2.3. The simulation assumes that each long-haul flight is assigned an arrival fix and a CTA by the integer program. The flight then proceeds to the metering fix and attempts to arrive at the assigned CTA. Additional uncertainty was applied to the travel times between the boundary and the metering fix so that flights arrive within the vicinity of their CTAs but not necessarily at that specific time. Short-haul flights were randomized based on samples from an empirical CDF based on historical departure delay data for ATL over the month of May, 2011. The short-haul unmanaged flights were merged with the CTA assigned flights outside 500 nm to create an integrated stream of flights based on the arrival times adjusted after randomization. The simulation then processed the flights into vacant slots according to a first-in, first-out (FIFO) queuing process. When the demand for the runway space exceeded capacity, the flights were held and the resulting delay was measured. A baseline run was used to evaluate the delay performance with no intervention. This trial used flight ETAs and projected them backward to get the approximate arrival time at the metering fix. Once the baseline run was completed each model was tested under the same simulation conditions.

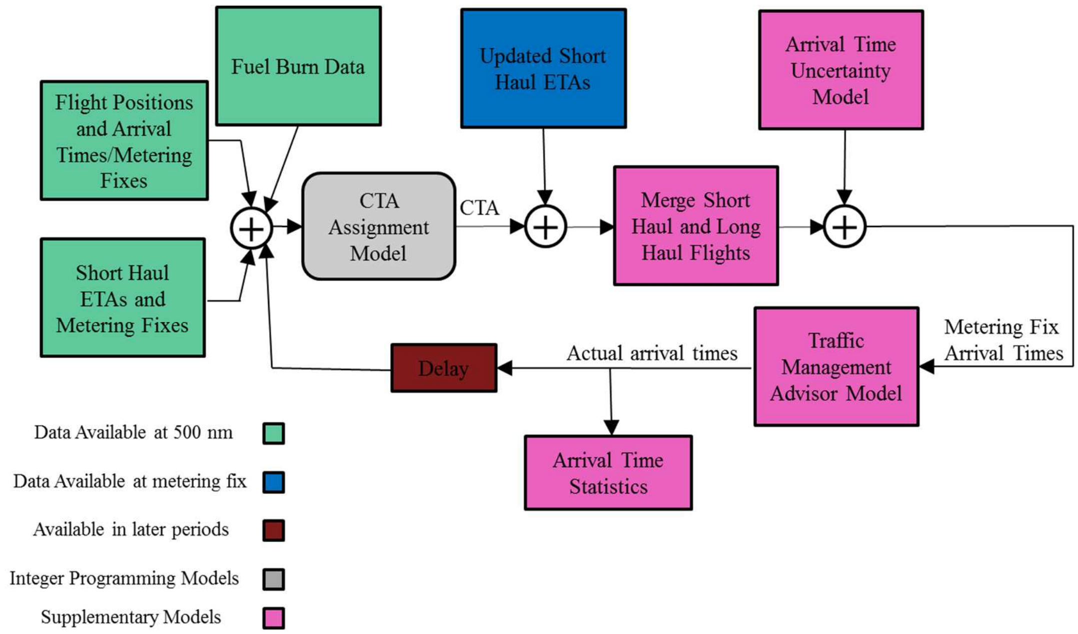

Figure 2.3 An illustration of the Model Simulation Framework.

2.3.2 Generating the Uncertainty

We developed a distribution of fix-to-runway flight times by sampling those data from the historical record for this airport on May 1, 2011. Those data should be equally applicable whether arrival time controls are in effect or not. The departure delay distributions for short-haul fights were derived using historical departure delays for all airports serving Atlanta during the month of May 2008. It turns out that no Ground Delay Programs were executed during this month at Atlanta. This is important because we suspect that departure delays for short distance flights under a GDP would have significantly less variance than if no GDP were run. It should be cautioned, therefore, when conducting experiments such as these, to take care not to mix data from GDP days and non-GDP days if doing so would bias the distribution of delays.

These data were first used to generate the scenarios and associated probabilities for the stochastic IP. The same distributions were used in the monte carlo simulation test environment to evaluate the solutions. In the case of the FA model the continuous delay distribution was used to generate the expected number of short-haul flights in each slot. We assumed that the distribution was centered at the ETA of each flight. A density function was generated from the samples and the probability of a flight landing in each given slot was calculated by summing between the appropriate time intervals. These probabilities were aggregated to calculate the expected number (rounded to an integer) of flights in a given slot.

For the scenario-based model an empirical CDF was used to calculate the slot arrival times. In each scenario a sample was taken from our distribution and used to calculate the deviation from the ETA. This deviation was then added to the ETA to place the flight in the appropriate slot. This process was repeated until all ETAs were appropriately adjusted. This distribution was also used to generate the uncertainty in the simulation environment. Samples were collected from each short-haul flight and were added to the ETA.

In addition to the large source of uncertainty originating from the variation in short-haul departure times, we also accounted for the variation in travel times from the metering fixes to the runways. A normal distribution centered at the ETA was generated by sampling from historical data. These samples were tailored to each metering fix. The simulation assumes that each flight uses the same cornerpost it flew to in the historical data.

2.3.3 Measuring the Delay Savings

A simulation of the whole system was run to evaluate the ability of our model to transfer delay away from the terminal. A baseline measurement was performed to gauge the amount of delay present in the terminal without our intervention. In the set of runs constituting the baseline measurement, we recorded the amount of time that each flight spent in terminal airspace while waiting for a runway. If a flight arrived in the queue and could not receive a runway slot when it was within the allotted travel time, it then waited until a space opened up. This waiting time was measured and averaged. We then configured our model to assign CTAs to flights near 500 nm of the airport in 15 minute intervals using the IP model under test. We repeated the runs with the assigned CTAs and measured the average delay per flight. This delay was compared to the average delay without intervention to calculate the delay savings. For clarity, an expression for calculating transferred delay is shown in equation (2.15).

$$Avg\_D_{transferred} = Avg\_D^{baseline} - Avg\_D^{CTA} \quad (2.15)$$

Figure 2.4 shows an example of the delay curves yielded by the model along with the resulting delay transfer. The solid curve reflects the delay accrued with no model intervention. The dot-dash curve reflects the portion of the delay that was transferred away from the terminal. The gap between the two curves shows the residual delay left in the terminal area after model intervention. Note that while the model does not reduce the overall minutes of delay per flight, it significantly altered the level of delay absorbed during the cruise and terminal phases of flight.

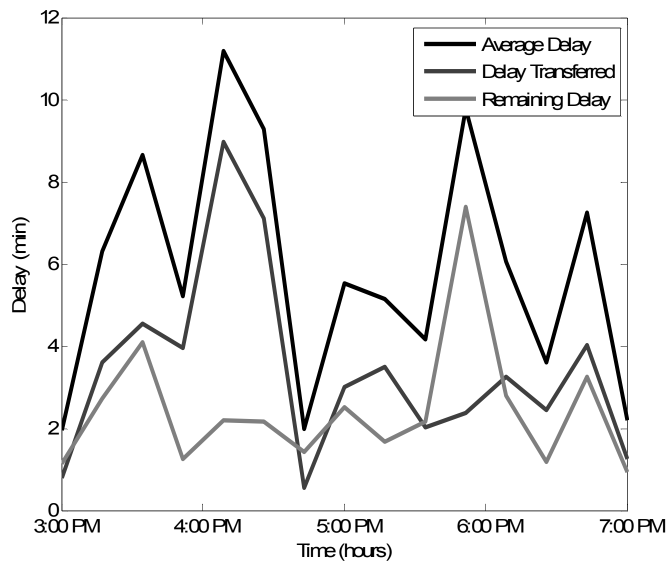

Figure 2.4 An example of the delay and delay transfer levels over a 4-hour period.

2.3.4 Model Performance

Each of our three models was tested in our simulation environment using data collected over a 4-hour period. Figure 2.5 shows the delay transfer of our models relative to the total delay. The figure suggests that all of our models show some ability to transfer delay away from the terminal. All delay transfer curves mimic the shape of the terminal delay curve. The deterministic model was ineffective, transferring 3.83% of the delay. This indicates model actually adds delay to the flights on an aggregate basis. We did, however, find substantial improvement when we attempted to account for the demand uncertainty using our stochastic models. The Functional Approximation model transferred 19.17% of the delay. The delay transfer resulting from the scenario-based approach ranged from 12.58% to 19.53% based on the number of scenarios used.

In order to make the comparisons most meaningful, we used the same random number seed for any given iteration of the three models. This ensured that the underlying realization of the flight delays was common between the three optimization models. Any difference in performance, therefore, is related only to the behavior of the optimization models, and not to a happenstance of the random variables. Different iterations of the simulation, of course, used different and independent random number seeds.

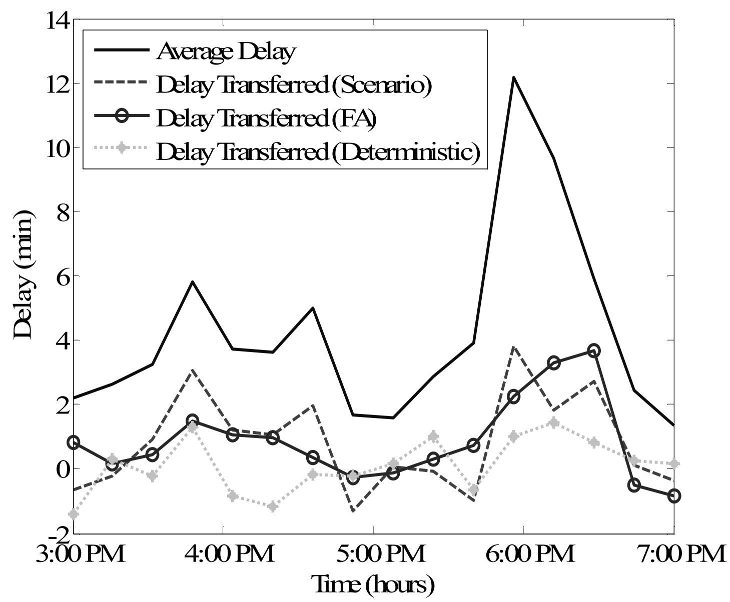

Figure 2.5 The delay transfer performance for all three models over a 4-hour period.

The upper bound of the resulting delay transfer from the scenario-based approach exceeds that of the Functional Approximation model. However, this comes at a significant cost in computation time. To understand the full extent of the performance we tested the computation run time of each model using a dual-core system with four Intel Xeon X5535 processors and 12 GB of memory in a 64 bit environment. The models were coded in Python 2.7 using a GUROBI solver. Each test case was generated using 100 trials. The results of the runs are shown in Table 2.1.

Table 2.1 A Summary of Model Performance

Model	Number of Variables	Number of Constraints	Mean Delay Transfer	Std Dev. Delay Transfer	Average Solution Time (seconds)	Std. Dev.Solution Time (seconds)
Deterministic	191	386	-3.8266	9.6163	0.1022	0.0277
FA	291	628	19.1679	7.2377	0.0904	0.0196
Scen 100	5643	5676	12.5843	7.3585	0.6853	0.0722
Scen 500	27243	26076	18.0990	7.3459	4.1022	0.3081
Scen 750	40743	38826	18.6516	6.6373	7.5442	0.7264
Scen 1000	54243	51576	19.4714	6.7288	9.5569	1.1395
Scen 1500	81243	77006	19.5005	6.6536	13.9108	1.0639
Scen 2500	135243	128076	19.5254	5.6868	24.9520	2.0245
Scen 5000	270243	255576	19.0537	6.1001	64.1577	8.0384

The table shows that when a small number of scenarios is used, the scenario-based model cannot account for the uncertainty well enough to match the delay transfer performance achieved by the FA model when the number of samples is small. When a larger number of scenarios is used to model demand uncertainty the delay performance exceeds that of the FA model; however, it does so at a significant computational cost. In the instance of the 1000 scenario test case the model runs 2 orders of magnitude slower than the FA model. If we needed to add additional scenarios to account for the effect of capacity uncertainty or attempted to extend the model over multiple airports, thereby increasing the problem size, this would only compound the problem from an implementation standpoint. Thus the FA model proves a stronger candidate to deal with the various facets of the problem. It should be noted that, for the third model, scenario generation must take place whenever a problem instance is solved, e.g. at each 15-minute time interval in the rolling-horizon implementation. Further scenario and problem instance generation can take a considerable amount of time (more than IP solution time) when a larger number of scenarios is used. Since we did not use a particularly efficient means for implementing this step, we do not report those results here.

2.3.5 Fuel Burn Savings

While we have focused on the mechanics of transferring delay away from the terminal, our primary objective is to save fuel. We would like to understand how such delay savings translates into fuel conservation. In order to measure the average fuel savings we needed to conceptualize how the savings occurs. Transferring delay on a given flight from the terminal area to the en route phase of flight can save fuel. While some of this savings results from transferring the site of the delay from a lower to a higher altitude, the majority of the benefit is attributable to the reduction in distance traveled. As we discussed in section 1.3.1, terminal delay is applied largely by extending the paths of flights. By transferring the delay to the en route airspace we are able to eliminate a considerable portion of the extended path. Since the fuel burn rates en route are nearly equivalent for the standard and speed controlled flights, the conservation of fuel achieved through the reduction in path extension in terminal airspace is essentially free.

In order to explicitly calculate the average savings rate incurred on a per flight basis, we measured the fuel burn rate near the terminal at various altitudes. We assumed that the aircraft vectoring inside the terminal would do so at altitudes over a range of FL100 to FL250. With this range we sampled altitudes from an empirical inverse CDF derived from flight trajectories in the terminal airspace of ATL. These altitudes were then used to measure the average fuel burn rate at a given speed based on values obtained from the BADA (Base of Aircraft Data) database (Eurocontrol, 2014). The results of these computations can be seen in Figure 2.6 below.

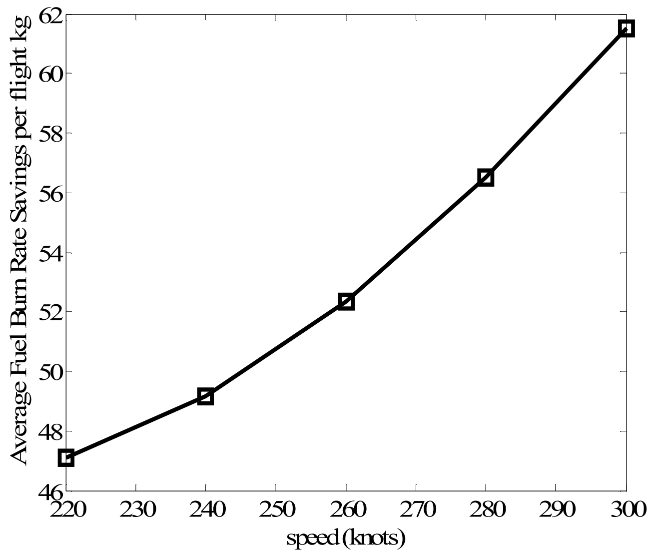

Figure 2.6 Average Fuel Burn Rates Savings Rate (kg) from the total fleet mix vs. Speed (CAS).

Given the inherent fuel burn savings rate associated with moving small amounts of delay away from the terminal it is illustrative to examine how the delay transfer curve translates directly into fuel savings. Figure 2.7 shows the fuel burn savings made possible by the delay transfer relative to five different vectoring speeds at the terminal. A comparison of the plots shows that the savings is considerable regardless of vectoring speed. Although the savings is largest when vectoring at 300 knots, in every case in this example we are able to save an average of 54.65 kg per flight over the 4=-hour period.

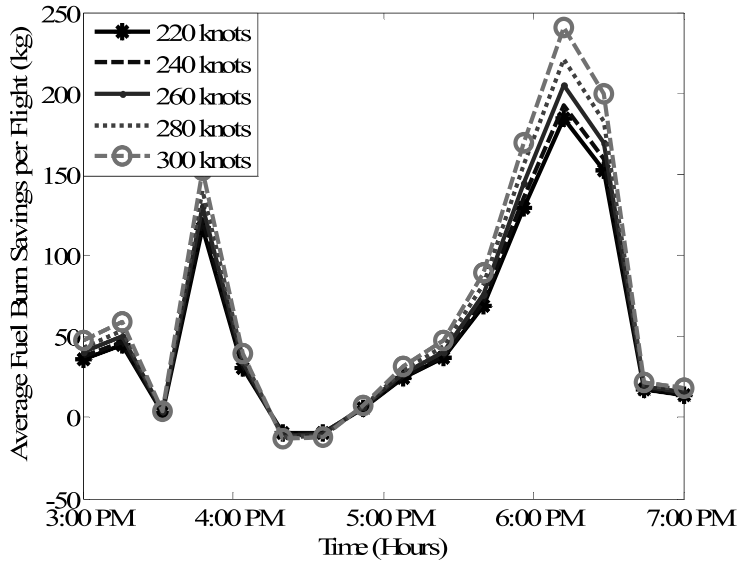

Figure 2.7 Average Fuel Burn Savings per flight vs. time over a 4-hour period using the Functional Approximation model.

2.4 Managing Flights with Restrictions

In the previous two sub-sections we assumed that flights will make a good faith effort to arrive at their CTAs. There are, however, a number of operational constraints that may inhibit some flights from even attempting to meet their prescribed arrival times. Flights do not operate in a vacuum. The arrival flows of flights operate in streams insomuch as the aircraft generally follow each other on a successive path. As a result, when the arrival time of one flight is changed, it can alter the arrival times of a number of other successive flights. Since these flights are often not going to the same airport and therefore not coordinated together, assigning to one flight a CTA that is substantially different than its original ETA could negatively impact the ability of the other flight to arrive at its ETA. For this reason it may be impractical in certain instances to compel a flight to meet its CTA.

In addition, flight operators have other operational constraints that may influence their desire to comply with CTA assignments. For example if a flight has a number of passengers on a connecting flight leaving shortly after the ETA and it is assigned a CTA which effectively delays the flight, it may wish to ignore the CTA and arrive at its ETA. While this behavior is not desireable and one would ultimately aim to work with air carriers to better suit their preferences, under this proposed set of models it is not unreasonable to expect some flights to deliberately disregard their CTAs and arrive near their ETAs. As a side note, this is exactly the kind of situation and behavior that a Collaborative Decision Making (CDM) implementation of these ideas would seek to prevent. In this section we explore the implications of these two scenarios. We present a revised functional approximation model and test its performance in the presence of restrictions on the set of allowable CTAs.

2.4.1 Functional Approximation Model with Restrictions

The prior two functional approximation models assumed that flights were capable of arriving over a range of CTAs that were governed by the set of allowable speeds of the aircraft. In this model we assume that some flights must not deviate from their ETAs. When this occurs the set of arrival time restrictions is known prior to the 500 nm boundary and this information can be incorporated into the assignment process. Accordingly, flights that must arrive at their ETAs are restricted to the CTA slots nearest their ETAs. Flights that have no restrictions are free to be assigned CTAs that can be met based on the performance of the aircraft. To simplify our test case and to isolate the effect of this modification, we reverted back to our assumption that managed flights will meet their CTAs with certainty.

To incorporate the appropriate CTA restrictions, the model assumes that a pool of flights is randomly chosen from a set of unmanaged flights on each iteration. Once this pool is identified, these flights are then restricted to the slots nearest to their ETAs. A description of our slot restriction model is given below:

Additional Parameters

V_{if}- ≡ The set of all slots available to flight f at fix i when flight f has restricted movement

P≡ The set of flights restricted to their ETA

$\mathbb{Z}^{+}$≡ The set of positive integers

$$\min \sum_{s\in Y_r, r\in R} W^{sr} - Mv \qquad (2.16)$$

$$\text{s.t.}(2.2-2.5),(2.8),(2.10),(2.14),(2.15)$$

$$\sum_{\substack{k\in V_{if}, i\in\Omega_f, \\ s\in Y_{rf}, r\in R}} x_{fks}^{ir1} + x_{fks}^{ir2} = 1 \quad \forall f \in P \qquad (2.17)$$

$$\sum_{\substack{f\in F, \in Y_{rf}, \\ r\in R}} x_{fks}^{irj} \leq 1+v \quad \forall k \in S_{if}, i \in \Omega_f, j \in T \quad (2.18)$$

$$\sum_{\substack{f\in F, k\in S_{if}, \\ i\in\Omega_f,}} x_{fks}^{irj} \leq 1+v \quad \forall s \in Y_{rf}, r \in R, j \in T \quad (2.19)$$

$$v \in \mathbf{Z}^{+} \qquad (2.20)$$

As in our first two functional approximation models our objective minimizes the total queueing delay. In equation (2.17) we introduce a set of constraints that forces flights in the restricted pool of flights to adhere to their ETAs. The additional constraints introduced due the added flight restrictions rendered the problem infeasible in a few instances. To account for this complication we introduced a relaxation by adding a variable v to equations (2.3) and (2.4).

2.4.2 Performance with Restrictions and Limited Compliance

To evaluate the performance of our interventions we performed two computational experiments. In the first we repeated the experiment described in section 2.3.4 using a pool of restricted managed flights. The percentage of managed restricted flights was varied from 0% to 30%. The uncertainty on all flights entering the terminal was generated using the same normal distribution as in section 2.3.4. A plot of the resulting average delay transfer over the 4-hour period versus the percentage of restricted flights is shown in Figure 2.8. The performance remains relatively consistent when the percentage of managed unrestricted flights ranges between 80% and 100%. When the percentage of unrestricted flights drops to 75% there is a noticeable decline in performance. This drop suggests that the level of intervention on the part of air traffic controllers could have a noticeable impact on performance. Even with this drop, however, the model performance still exceeds 8% of the total delay, suggesting that the model can transfer modest amounts of delay in a restrictive environment.

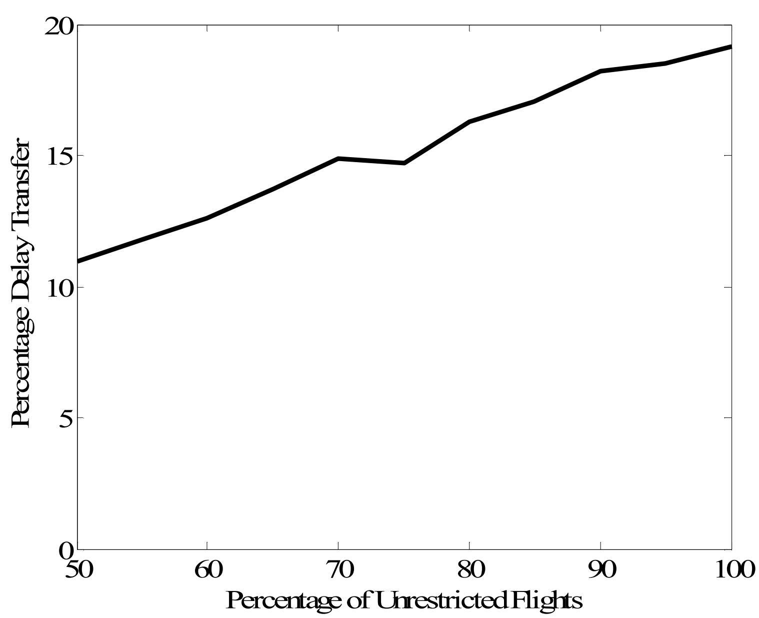

Figure 2.8 The Percentage Delay Transfer vs. Percentage of Unrestricted Flights.

An additional experiment was conducted to evaluate the performance when some flights choose to willingly disregard assigned CTAs to meet their internal objectives. In this test case flights were assigned CTAs using the functional approximation model presented in section 2.2.4. We then identified a certain percentage of these flights and moved their arrival times to their ETAs. Uncertainty near the terminal was added to the flights using the normal distribution employed in our previous experiments. In this experiment the percentage of compliance was varied from 100% to 30%. Figure 2.9 presents the mean delay transfer over the 4-hour period versus the percentage of restricted flights. As expected, the system performance increases with compliance. The precipitous drop at 75%, suggests that like the variation in the number of restrictions there is a threshold level below which the model exhibits a substantial and dramatic decline. At 35% the performance drops to levels at which the intervention is ineffective, however, the stability up to this point suggests that the model is able to transfer a considerable amount of delay, even in the presence of less than ideal compliance.

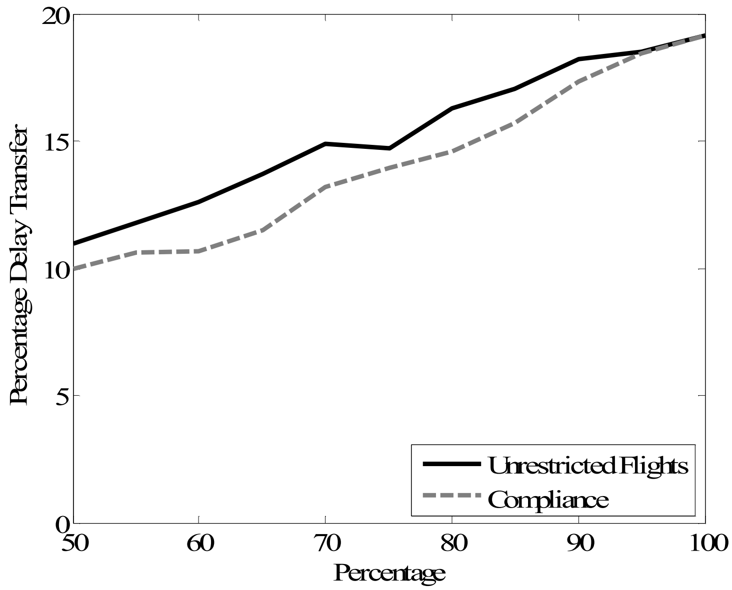

Figure 2.9 The Percentage Delay Transfer vs. Level of Compliance.

2.5 Sensitivity Analysis

The analysis thus far was performed using data taken on a single day at one airport over a 4-hour period. While the proposed methods have shown noticeable improvement under these circumstances, the effectiveness could be more conclusive if tested under a broader range of conditions. There are a number of ways in which the conditions could differ from those tested. The airport could be operating at reduced capacity, the estimated arrival times could differ and the departure delay distribution of short-haul flights could vary. To address these issues we performed three additional tests to evaluate our model performance.

In our first test case we examined the effect of the departure delay distributions of the short-haul flights on delay transfer. We examined two additional distributions both of which are delayed relative to the STAs of the short-haul flights. A plot of all three distributions is shown in Figure 2.10. The delay distribution generated based on the May 2008, Atlanta airport data sample indicates that in the absence of TMIs the vast majority of flights depart prior to their STAs. While this may be largely true it is conceivable that in some instances flights may depart later than their STAs. To test the impact of these later departures we applied gamma distributions such that all flights departed later than the STA. In each instance the functional approximation model was fitted with the distribution of short-haul flight departure times.

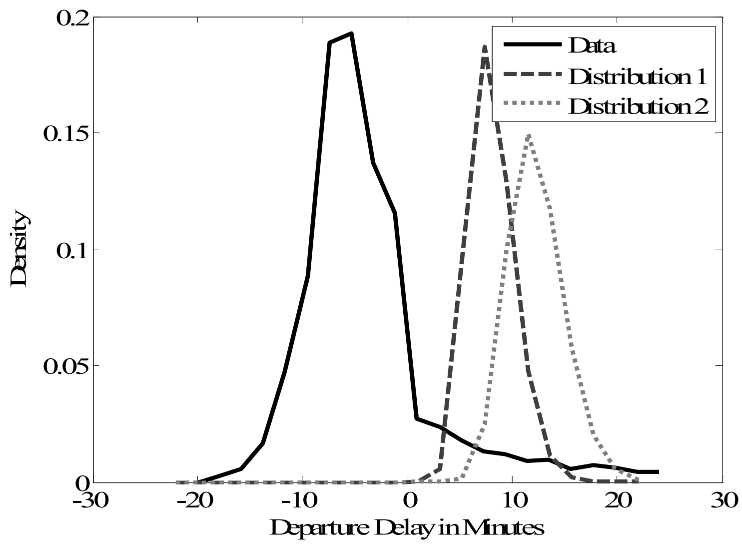

Figure 2.10 Departure delay distributions for short-haul flights.

It is also possible that the airport will operate at a reduced capacity in some instances. If this capacity is sufficiently large, Traffic Flow Managers may seek to implement a GDP. In other situations Traffic Managers may wish to use both mechanisms in concert. While still in other instances the capacity may be substantial yet not at the level at which a GDP is necessary. To gauge the effect in situations in which the capacity of the airport is slightly more compromised, we explored two test cases where the capacity was reduced to increasing levels. Since there is no assumed ground delay applied to the flights the impact of this additional capacity must be absorbed through increased airborne holding. Thus the pool of delay available is significantly larger. To deal with this issue the slot size in the model was adjusted to match the reduced capacity of the airport.

It is quite likely that the scheduled and estimated arrival times will differ from those seen in the single day of data we examined. In some instances the STAs of flights may clump more closely together to create larger peaks in demand while in other instances the demand may be more evenly spread. To investigate the impact of STAs on our model we randomized the arrival times from our baseline sample. A sample from a uniform distribution with maximum and minimum times of +/-10 minutes was used to perturb the data.

For each of the test cases described the simulation was run 100 times. A summary of our results is shown in Table 2.2.

Table 2.2 Percentage of delay transfer under alternative conditions

Test Case	mean	std
Original Distribution	19.1679	7.2377
Distribution 1	25.61968097	6.79249189
Distribution 2	24.79977414	7.414530151
10% Capacity Reduction	30.16903728	6.043011608
5% Capacity Reduction	28.19081126	6.546273671
ETA1	23.44633691	6.916352094
ETA2	16.29709966	5.828945401
ETA3	18.23657348	5.785250686
ETA4	18.20156288	6.811964605
ETA5	18.48333541	7.020638307

The table suggests that delay transfer performance should improve in the presence of the alternate short-haul flight delay distributions. This follows from the fact that in our original distribution flights are biased to arrive ahead of their ETAs. Since these flight are entering an arrival stream that is already backed-up they need to wait longer before the airport can accommodate them. Thus when this assumption is relaxed flights experience less airborne delay. The capacity reduction also acts to improve the amount of delay transferred. In this case the reduced capacity of the airport induces longer delays. This allows our model to operate more effectively because the opportunity for delay transfer relative to the uncertainty in the arrival times is greater. Variation in ETAs induced some variability in the percentage of delay transfer but did not prove a significant contributor to altering the gain experienced from the use of our model. Overall, the experiments lend increasing evidence to our model's ability to perform effectively under a wider range of conditions.

2.6 Conclusions

In this chapter, we presented three models to transfer aircraft delay away from the terminal airspace. The first was a deterministic model that sought to maximize throughput, and that serves primarily as a baseline against which to measure more realistic stochastic variants. The second model was stochastic and incorporated scenarios to account for assumed demand uncertainty. A third model used a functional approximation of the expected number of flights to minimize the expected excess delay in the terminal area. While all approaches demonstrated an ability to transfer delay from the terminal area to the en route airspace, the two stochastic models proved more effective. An analysis of the computational performance of each model showed that the functional approximation model demonstrated efficient run times relative to higher fidelity scenario models while achieving comparable delay transfer. This translated into significant fuel savings on a per flight basis when fuel burn was analyzed.

The chapter also analyzed the performance of our functional approximation model under a set of declining operational conditions. In some instances the model was adapted to better suit the changes to the operational environment. When uncertainty was introduced into managed flights, a revised model was able to achieve performance at a level comparable to that achieved when there was no uncertainty introduced on managed flights. When flight restrictions were introduced, the model was able to perform at a comparable level until the number of restricted flights reached 25% of the total flow. Even when the compliance level dropped, the model demonstrated substantial delay transfer. The resilience of the model suggests that it could prove a strong candidate to achieve delay transfer.

The work in this chapter presents a couple of promising opportunities for future research. Along with demand uncertainty, capacity uncertainty also poses significant challenges in its effect on holding within the terminal. The efficient run times of the functional approximation model indicate that it could be extended to deal with both types of uncertainty even if additional scenarios were used to model capacity. Furthermore, the model could also prove effective in a multi-airport setting in which more resources are utilized. Finally, in this paper we have made the assumption that the intervening ANSP has no active control over the short-haul flights. A version of the model that assumes greater control over short-haul flights could also be studied.

3 Transferring Delay through Control of Airborne and Ground-based Flights

In the previous chapter, we proposed a set of integer programming models designed to transfer the delay away from the terminal. One of the most challenging aspects of that problem is that we did not know when a substantial portion of the flights were going to arrive and assumed no control over their operation. While this may be a realistic constraint, it is also certainly possible to envision an environment in which the ANSP (FAA) assumes a more active role in coordinating all flights. While issuing speed changes to a short-haul flight will do little in and of itself to delay its arrival time, it may be possible to assume additional control through the use of ground delay assignments. In such a system, short-haul flights could be assigned a combination of ground delays and relatively small airborne delays, while long-haul flights could continue to be assigned airborne delays through en route speed control. In this chapter, we explore the concept and propose three new stochastic integer programming models to deal with this revised set of controls and uncertainty. In section 3.1 we present our conceptual revisions to the problem. In section 3.2 we provide a description of the model. In section 3.3 we discuss our experimental results. In section 3.4 we provide some concluding remarks and suggest extensions to the work presented.

3.1 Conceptual Revisions

In the previous chapter, we stated that our goal was transfer the delay from the terminal to the en route phase of flight. We showed that by assigning delays to flights en route, a considerable portion of the delay could be moved. While this delay transfer largely benefits all flights involved, the extent of the benefits is not equally realized. The models presented assign CTA to the pool of flights outside of a fixed radius. The remaining flights inside of the radius are allowed to proceed unimpeded towards their destination, yet their time spent in airborne holding is reduced.

If the pool of flights were equally distributed amongst the carriers that use that airport, one could argue that they were all being treated fairly. In reality, however, the pool of short-haul flights favors regional carriers, which puts them in a slightly better position. It is considerably more invasive to issue delays to control flights that have not left the ground due to the uncertainty associated with their departure times, however, it is still possible to delay these grounded flights on an aggregate basis while accounting for variation in departure time. While the pilots and crew may need to be given some notice in order to adhere to the ground delays, since the magnitude of the delays is significantly smaller than the type of delays typically taken during GDPs, the impact on gate occupancy should not be nearly as severe. Eventually such instructions could be given to crew over datalink, however, in the nearer term one might utilize tower control or airline dispatch to provide guidelines. A revised overview of our information transfer process is shown in Figure 3.1 and Figure 3.2.

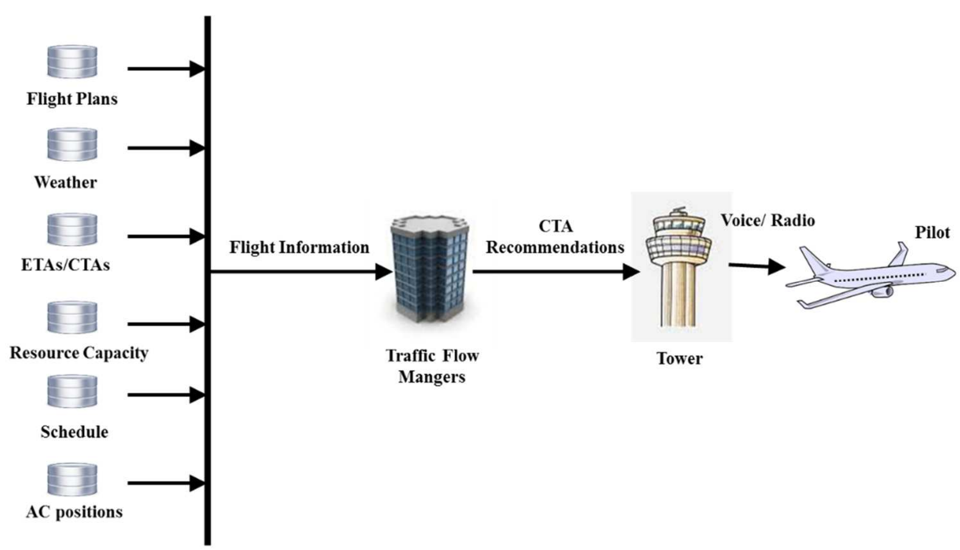

Figure 3.1 Information Flow under a Central Control Architecture

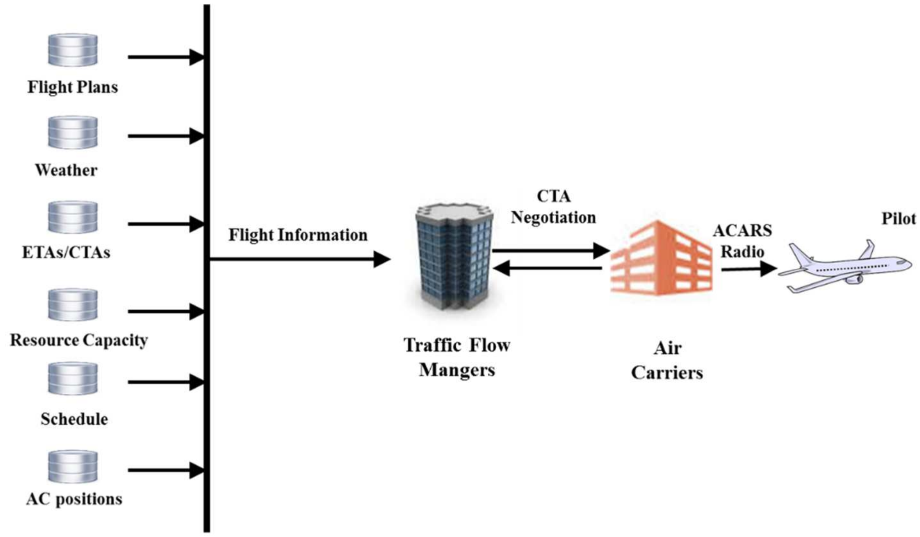

Figure 3.2 Information flow under a Collaborative Architecture.

One potential issue with incorporating ground delay intervention is that holding flights on the ground could have significant adverse impact on inbound flights and induce significant additional runway holding at the originating airport. If this concept were to be implemented, one would need to thoroughly understand the ability of airports to accommodate widespread ground delays in a network setting. If the scope of intervention were limited to a couple of destination airports, the effect would likely be marginal; however, were the concept to be extended over several airports, stakeholders may have to begin to prioritize flight delays at the origin or destination airports. One means of dealing with this issue is to incorporate collaborative decision making into the planning process. While a comprehensive examination of such questions lies beyond the scope of this dissertation we do attempt address some of these issues in chapter 4. In this chapter, however, we will consider the effect of applying ground delays in a single-airport context ignoring the larger effects on the network.

Operationally, there are a number of open questions regarding how ground and airborne delay should be administered. Since flights originate from different distances, they will have varying degrees of uncertainty associated with their travel times. Typically, flights that are in the air when a CTA is assigned will be more able to meet that CTA by virtue of the fact that they do not have to deal with potential departure delays. Additionally, CTA that are assigned to airborne flights at distances further away from the airport are less likely to be adhered to than CTA assigned to flights closer to the airport. Typically, GDPs address this problem by assigning capacity to carriers through slots of uniform width. Due the varying degrees of uncertainty associate with ability of the flight to meet its arrival time, some flights will be able to arrive during the slot window while others will deviate. When the deviations occur, flights assigned to one window may arrive in another thereby causing airborne holding. The resulting airborne queueing delay can be reduced by incorporating stochastic models of the queuing delay in the decision making process such as those described in the previous chapter. One drawback to this approach, however, is that such decisions are made hours in advance of the time the flights reach the runway. As such, our intervention may be overly aggressive or conservative depending on when the flights actually arrive. Such situations are likely to become more apparent between the time that the decision is made and the time the flight reaches terminal airspace. In these situations, it may be appropriate to issue an initial assignment at an earlier time and issue a recourse decision as flights get closer to the airport and more information becomes apparent. In the presence of this diminished uncertainty, it may be preferable to adopt a more opportunistic throughput oriented approach. An example of this approach is depicted in Figure 3.3.

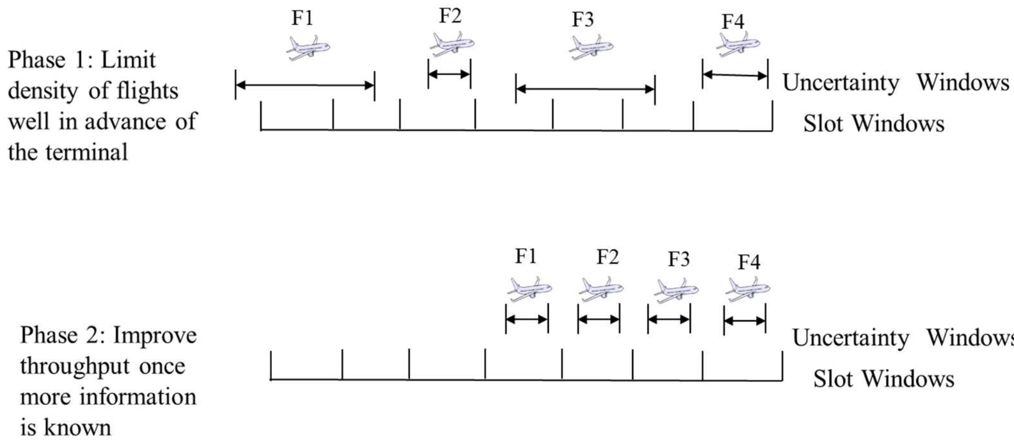

Figure 3.3 The proposed strategy for issuing CTAs to flights.

In this chapter, we investigate the viability of this strategy. In the next section, we present two stochastic models that could be used to assign the initial CTAs to flights. In the following section, we pair these models with the deterministic throughput model and examine the effect of using the two models in series.

3.2 Methodology

In this section, we present three stochastic integer programming models designed to achieve comprehensive control over all arriving flights. As in the previous chapter, we describe one scenario-based model and two models which relies on a functional approximation. The first builds off the scenario-based model introduced in the previous chapter by incorporating ground and airborne flight into the set of controllable flights. The model also adopts the Bertsimas "by" variables which have been shown to perform well in other integer programming ATFM problems. The second and third model each track probability of deviation from the assign time and use this information to reduce the airborne queueing delay.

3.2.1 Scenario-Based Air-Ground Speed Control

The first approach considered in this chapter is a scenario-based model that assigns arrival times to flights located in both the air and ground. In order to determine which flights will get delayed in the air and on the ground, we group the flights into three separate pools. Flights originating outside an outer-fixed radius receive airborne delay once they approach the radius. Flights originating inside some inner-radius will be issued solely ground delay. Flights originating between the inner and outer radii are issued a mixture of air and ground delay. An example of the notional access pools is shown in Figure 3.4.

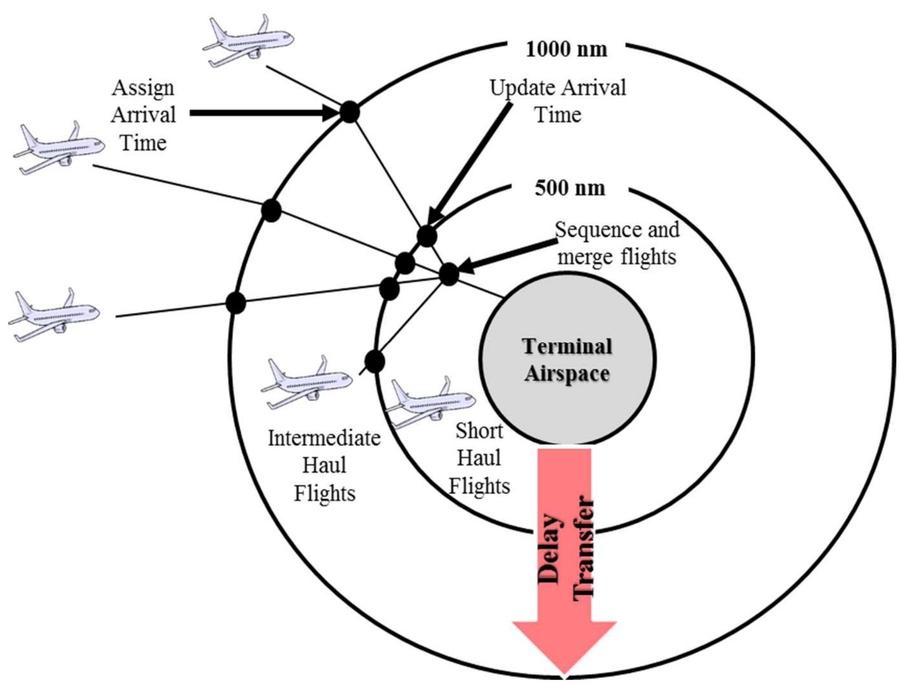

Figure 3.4 The partitioning of flights into assignment pools.

The objective of the model is to apportion access to a single airport while minimizing the total delay in the air and on the ground. As there is considerable amount of uncertainty associated with the ability of some flights to meet their arrival times, we utilize a stochastic programming model to manage the airport demand. The model assigns arrival times to flights into two stages. In the first stage flights are assigned an initial arrival time. This represents an initial CTA that would be assigned approximately two hours in advance of the arrival. While the pilot may make a good faith effort to ensure that he/she meet this CTA, there are a number of factors that could influence the outcome including: wind uncertainty, convective weather, departure delays and the availability of direct routing. Due to these and other factors, it is not unreasonable to conclude that there will be some deviation from the assigned arrival time. Since these times are issued well in advance of the flight's arrival, there is time to make a recourse decision and take corrective action. Our model maps the outcomes of the scenarios into its decision-making and uses them to issue the best recourse decision for each outcome. As we did not want to inject additional uncertainty into gate availability at the origin airport, we assumed that once a flight was delayed on the ground, the ground delay was fixed. Thus for a flight originating inside the inner radius, the initial delay would remain in place. For flights originating outside the inner radius, however, a recourse decision could be used to revise the CTA in order to reduce the amount of airborne holding due to conflicting arrival times.

When the decision is made, the set of time slots available for potential assignment are set based on a direct mapping of the feasible aircraft speed and the distance from origin to destination at the time of assignment. The recourse decision is also bound by this mapping. Thus a flight assigned to arrive 5 minutes ahead of schedule, for example, cannot be revised to arrive 15 minutes ahead of schedule if it cannot reach the destination while traveling at the fastest possible aircraft speed. It may, however arrive at time later than what could be achieved by traveling at the slowest possible aircraft speed. In these situations, it is assumed that the aircraft will fly as slow as possible and begin holding in the air once it nears the airport. In addition, flights originating within the inner radius could arrive no earlier than their scheduled time of arrival. Since flights within the inner radius cannot be adjusted at the time of recourse, we also require that the slots which are occupied by these short-haul flights may not be filled by airborne flights. The assignment process is depicted for three situations in Figure 3.5, Figure 3.6 and Figure 3.7.

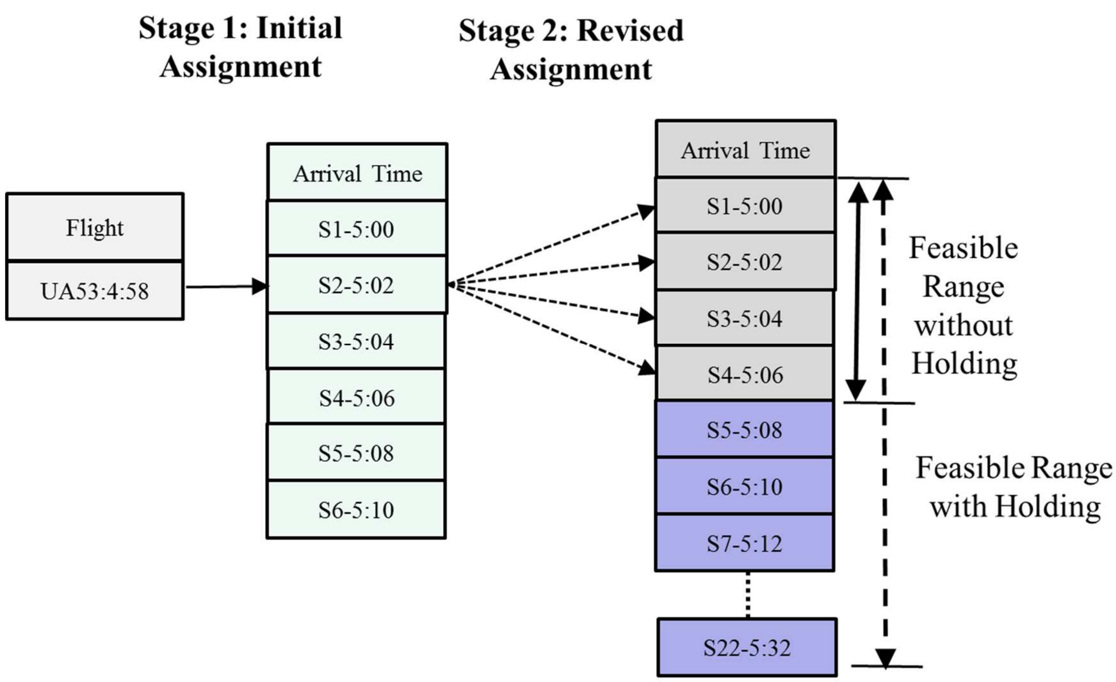

Figure 3.5 Flight assignment of an airborne flight over two stages.

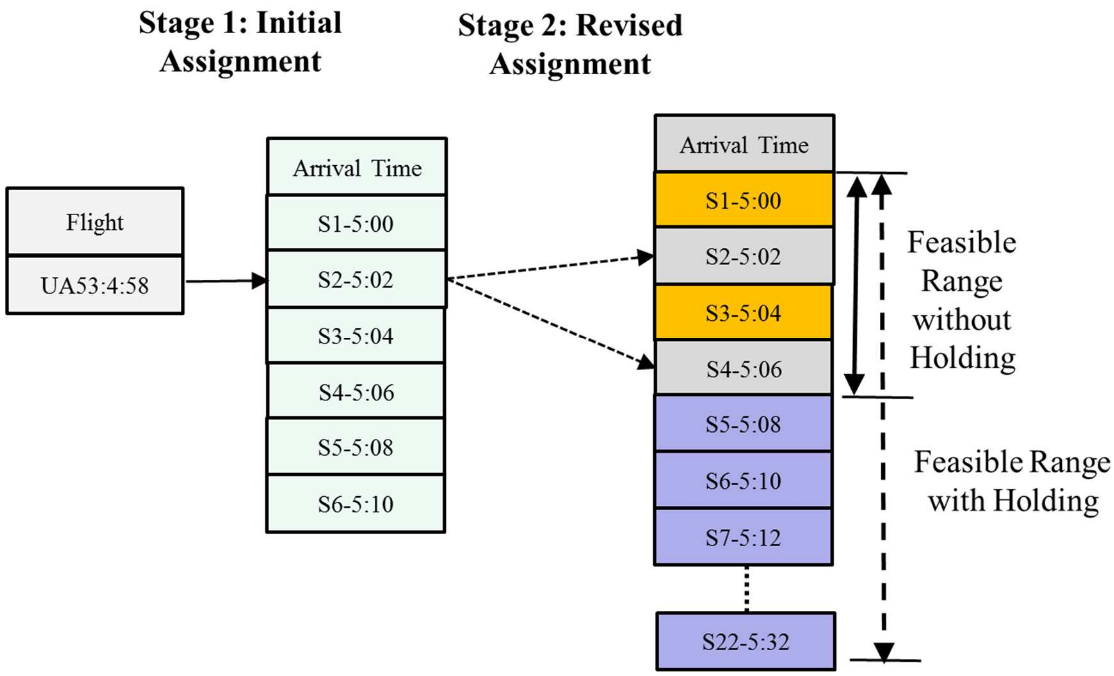

Figure 3.6 Flight assignment of an airborne flight over two stages with restrictions due to the presence of ground-based flights.

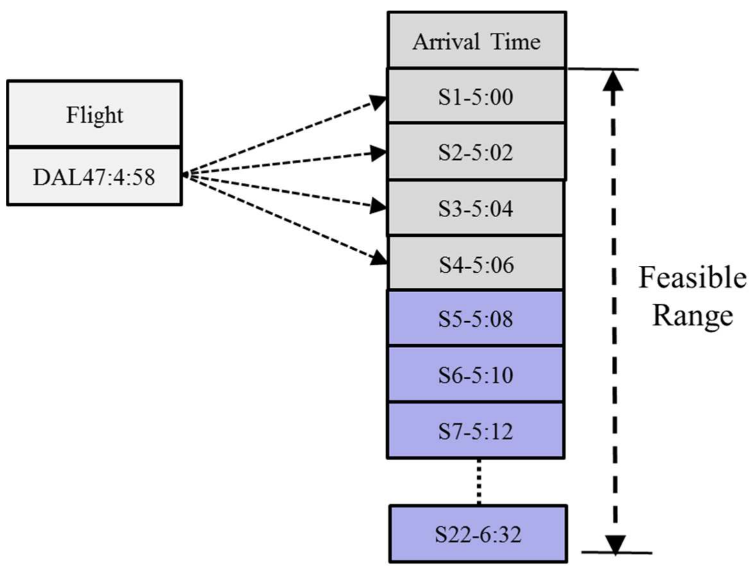

Figure 3.7 Flight assignment of a ground-based flight.

In each scenario, the model identifies the best slot from a set of reachable slots based on the initial CTA assignment using a weighted objective of air and ground delay. These components differ substantially in the way they are represented in our model. The ground delay realized by each flight is a function of both the assigned delay and the uncertainty in the departure time, as there is no means for recovering the ground delay at a later stage of the problem. Although we could incorporate the ground delay uncertainty directly into this portion of the objective function using scenarios, we also account for it in terms of its effect on the airborne holding. Thus to avoid double-counting the effect, we therefore use only the deterministic component of the ground delay to affect the objective. On the other hand, the airborne delay is assigned and reassigned over two stages. To deal with the reassignment, we use the sum of airborne delay over each scenario to estimate the queuing delay within the system.

For each scenario, there is a probability p_q associated with the probability of that outcome occurring. If we had considerable information about the environment, it might be possible to incorporate factors such as the variation in departure delay of similar days, the likelihood of weather and wind impacting specific fights routes and the level of congestion and time-varying capacity of various fixes into these probabilities. In our case, we have little information specific to the day of operation so we shall assume uniform probabilities.

We would like to make assignments to flights roughly two hours prior to arrival. Due to the time scale, it may be considerably more difficult to make assignment decisions that are specific to airport runways. As such, we chose to treat the system as a single server queuing system. The length of the queue will be the sum of delay carried over from the previous period plus the number of flights arriving in a slot minus the capacity of the slot. For the purposes of this chapter, we shall assume a time slot capacity of 1. A formulation of our stochastic integer program is shown below.

Parameters

$F \equiv$ The set of all flights

$V \equiv$ The set of all flights originating inside the outer radius

$S \equiv$ The set of all slots

$C \equiv$ The capacity of each slot

$Y_f \equiv$ The set of all slots available to flight f

$S_{ftq} \equiv$ The set of all slots available to flight f upon reassignment given it is initially assigned to arrive by time t under scenario q. Note that for flights inside the inner radius the set is restricted to 1 slot.

$\tau_f \equiv$ The assigned arrival time of flight f from slot t

$\lambda \equiv$ A coefficient used to weight the convex combination in the objective function

$toag_f \equiv$The scheduled arrival time flight f

$t_{sf} \equiv$The travel time flight of f in cruise at slot s

$p_q \equiv$ The probability of scenario q occurring

Variables

$$x_{ft} = \begin{cases} 1 \text{ if flight } f \text{ is assigned to arrive by time } t \\ 0 \text{ otherwise} \end{cases}$$

$$w_{fs}^{q} = \begin{cases} 1 \text{ if flight } f \text{ arrives by time } s \text{ in scenario } q \\ 0 \text{ otherwise} \end{cases}$$

$$y_{fs}^{q} = \begin{cases} 1 \text{ if flight } f \text{ holds at time } s \text{ in scenario } q \\ 0 \text{ otherwise} \end{cases}$$

$$\min \quad \lambda \sum_{f \in V, t \in Y_f} \left(\tau_f - \tau_{oag_f}\right)\left(x_{ft} - x_{ft-1}\right) + (1-\lambda) \sum_{q \in Q} p_q \left(\sum_{s \in S} y_{qs} \right) \quad (3.1)$$

$$s.t. \qquad x_{fT_f} = 1 \quad \forall f \in F \quad (3.2)$$

$$\sum_{f \in F} \left(x_{ft} - x_{ft-1}\right) \le C \quad \forall t \in Y_f, f \in F \quad (3.3)$$

$$x_{ft-1} \le x_{ft} \quad \forall f \in F, t \in Y_f \quad (3.4)$$

$$x_{ft} \le \sum_{s \in S_{ftq}} w_{fs}^q \quad \forall f \in F, t \in Y_f, q \in Q \quad (3.5)$$

$$y_{qs} - y_{qs-1} - \sum_{f \in F} \left(w_{fs}^q - w_{fs-1}^q\right) + 1 \ge 0 \quad \forall s \in S, q \in Q \quad (3.6a)$$

$$W_{0q} \equiv \hat{W}_q \qquad \forall r \in R, q \in Q \quad (3.6b)$$

$$x_{ft}, w_{fs}^q = \{0,1\}, y_{qs} \ge 0, \forall f \in F, s \in S_{ftq}, q \in Q, t \in Y_f \quad (3.7)$$

Equation (3.2) states that every flight has been assigned to a time period. Equation (3.3) states that no time period can be filled by more than one flight. Equation (3.4) ensures that if a flight is assigned by the preceding time period, it is also assigned in the current time period. Equation (3.5) enforces connectivity between the two sets of variables and says that if a flight is assigned by a time period in the first stage it also must be assigned to a reachable slot in the second stage in each scenario. Equation (3.6) tracks the airborne hold in each time period for each scenario. Our objective function in (3.1) aims to minimize a convex combination of air and ground delay for all flights.

3.2.2 Air-Ground Control through Functional Approximation

In the previous section, we defined a scenario-based integer programming model that could ration the access to the airport by controlling air and ground delay. In the previous chapter, we saw that we could accomplish the same ends with a model that relies on a functional approximation. While our objective is slightly different, the challenge remains: can we achieve comparable accuracy at a reduced run time?

Unlike before, however, we now exert some level of control over the ground-based flights. Thus, our aim is to cohesively assign the all flights to time slots given the varying levels of uncertainty associated with each flight's ability to meet its CTA.

Our functional approximation of the expected queuing delay exhibits a similar structure to the one in the previous chapter. There is one notable difference, however, in the fact that the number of short-haul flights is now a set of decision variables, rather than parameters that is governed by the number of assigned flights arriving in each slot in each scenario. To model the uncertainty in our flight assignments, we used a triangular probability density function. We assume that, should a flight be assigned to a specific slot, the probability that it will arrive in a specific slot is governed by the density function and the slot where the flight was assigned. Our model formulation is shown below.

Parameters

$F \equiv$ The set of all flights

$Y \equiv$ The set of all slots

$N \equiv$ Maximum number of slots that can be assigned within the neighborhood of a fix

$a_{s'sf} \equiv$ The probability of flight f arriving in slot s' given that it was assigned to slot s.

$\lambda \equiv$ A coefficient used to weight the convex combination in the objective function

$t_{oag_f} \equiv$The scheduled arrival time flight f

$t_{sf} \equiv$The travel time of flight f in cruise at slot s

Additional Variables

$$x_{fs}^{k} = \begin{cases} 1 \text{ if flight } f \text{ is assigned to slot } s \text{ through fix } k \\ 0 \text{ otherwise} \end{cases}$$

$y_s \equiv$ The expected number of flights arriving in slot s

$W_s \equiv$ The number of flights arriving in slot s

$$\min \sum_{\substack{f\in F, s\in Y_f,\\ k\in\Omega_f}} \lambda\left(t_{sf} - t_{oagf}\right)x_{fs}^k + \sum_{s\in Y}(1-\lambda)W_s \qquad (3.9)$$

$$s.t. \sum_{s\in Y_f, k\in\Omega_f} x_{fs}^k = 1 \quad \forall f \in F \qquad (3.10)$$

$$\sum_{f\in F} x_{fs}^k \le C \quad \forall s \in Y_f, k \in \Omega_f \qquad (3.11)$$

$$\sum_{f\in F, s'\in S_s} x_{fs}^k \le N \quad \forall s \in Y_f, k \in \Omega_f \qquad (3.12)$$

$$\sum_{s\in Y_f, k\in\Omega_f} a_{s'sf} x_{fs}^k \le y_{s'} \quad \forall f \in F, s' \in Y \qquad (3.13)$$

$$W_s - W_{s-1} + 1 \ge y_s \quad \forall s \in Y \qquad (3.14a)$$

$$W_0 \equiv \hat{W} \qquad (3.14b)$$

$$x_{fs}^k, y_{s'} \in \{0,1\}, W_s \ge 0, \forall f \in F, s \in Y, k \in \Omega_f \qquad (3.15)$$

Equation (3.10) assures that every flight is assigned to one slot. Equation (3.11) ensures that each slot can be assigned to no more than one flight. Equation (3.12) forces the number of flights assigned within a neighborhood of a given slot at a fix to not exceed a threshold value *N*. Equation (3.13) dictates the value of our continuous variable used to track the number of flights expected to arrive in a given slot. Equations (3.14) tally the expected number of flights in the queue during each slot interval. The objective function in equation (3.9) features a term for both ground and airborne delay and tries to minimize the sum of the two values. The terms are weighted using a convex combination to reflect the relative cost of air and ground delay. As in the previous problem, the model is intended to operate in a dynamic context and is solved iteratively every 15 minutes.

One potential drawback to the model is that it does not explicitly consider fairness in the allocation process. While it is difficult to justify imposing fairness in airborne holding because it is hard to predict exactly what the controllers might do when they have to issue allocation, there is some precedent for enforcing fairness with respect to ground delay during GDPs and AFPs. By utilizing coefficients that grow super-linearly, we can impose increasing penalties for issuing larger ground delays. An alternative objective is shown in equation (3.16).

$$\min \sum_{\substack{f\in F, s\in Y_f,\\ k\in\Omega_f}} \lambda\left(t_{sf} - t_{oagf}\right)^{1+\varepsilon} x_{fs}^{k} + \sum_{s\in Y} (1 - \lambda)W_s \quad (3.16)$$

The first term in the objective function grows super-linearly and makes it increasingly costly to assign each unit of delay. Due the different growth rates of the two terms in the objective function, it is questionable whether the interaction will produce an assignment that will yield strong delay transfer; however, given the potential to better accommodate carriers with “fairer” assignments, we choose to examine the effect when integrated into the functional approximation model.

The formulation above presents a baseline model that has the potential to transfer delay using far fewer variables and constraints than almost any instance of a scenario-based model. Yet the formulation could potentially be strengthened by adopting the by-variables presented in the previous section. To that end, we present an alternative formulation design to improve upon the computation performance of functional approximation. We shall refer to this model as the FA-by model.

<u>Additional Parameters</u>

$T_f \equiv$ The final slot available to flight f for assignment

<u>Additional Variables:</u>

$$x_{fs}^{k} = \begin{cases} 1 \text{ if flight } f \text{ is assigned arrive by the timestamp of slot } s \text{ through fix } k \\ 0 \text{ otherwise} \end{cases}$$

$$\min \sum_{\substack{f\in F, s\in Y_f,\\ k\in\Omega_f}} \lambda\left(t_{sf} - t_{oagf}\right)\left(x_{fs}^k - x_{fs-1}^k\right) + \sum_{\mathrm{s}\in \mathrm{Y}} (1 - \lambda)\mathrm{W_s} \tag{3.17}$$

$$s.t. \quad x_{fT_f}^k = 1 \quad \forall f \in F, k \in \Omega_f \tag{3.18}$$

$$x_{fs-1}^k \le x_{fs}^k \qquad \forall f \in F, s \in Y_f, k \in \Omega_f \tag{3.19}$$

$$\sum_{f\in F}\left(x_{fs}^k - x_{fs-1}^k\right) \le C \quad \forall s \in Y_f, k \in \Omega_f \tag{3.20}$$

$$\sum_{f\in F, s'\in S_s}\left(x_{fs}^k - x_{fs-1}^k\right) + x_{f1}^k + x_{fT_f}^k \le N \quad \forall s \in Y_f, k \in \Omega_f \tag{3.21}$$

$$\sum_{s\in Y_f, k\in\Omega_f} a_{s'sf}\left(x_{fs} - x_{fs-1}\right) + a_{s'1T_f} x_{f1}^k + a_{s'sT_f} x_{fT_f}^k = y_{s'} \quad \forall f \in F, s' \in Y \tag{3.22}$$

$$W_s - W_{s-1} + 1 \ge y_s \quad \forall s \in Y \tag{3.23a}$$

$$W_0 \equiv \hat{W} \tag{3.23b}$$

$$x_{fs}^k, y_{s'} \in \{0,1\}, W_s \ge 0, \forall f \in F, s \in Y, k \in \Omega_f \tag{3.24}$$

Equation (3.18) states that every flight must be assigned by its last available slot. Equation (3.19) states if flight has been assigned by its previous slot it is also assigned by its subsequent slot. Equation (3.20) states the number of flights arriving at a slot cannot exceed the capacity of the slot. Equation (3.21) says that the sum the flights in the neighborhood of slot through a fix cannot exceed some threshold value. Equation (3.22) defines the expected number of flights in a slot. Equations (3.23) track the number of flights in the queue.

It is also conceivable that we may gain some benefit by adding the following valid inequalities formulation in (3.9-3.15):

$$W_s \ge W_{s-1} - 1 \quad \forall s \in Y \tag{3.25}$$

$$W_s \ge y_s - 1 \quad \forall s \in Y \tag{3.26}$$

The equations state that the number of flights in the queue during a given slot must exceed the greater of the number of flights occupying the queue of the previous slot minus 1 and number flights entering the queue minus 1. This relationship could be further generalized to include references from earlier time slots. In this form one might use the relationship:

$$W_s \geq W_{s-m} + \sum_{v=s-m}^{s-1} y_v - m \quad \forall s \in Y \quad (3.27)$$

$$W_s \geq \sum_{v=s-m}^{s} y_v - m \quad \forall s \in Y \quad (3.28)$$

Where m is the number of slots back that we wish to reference. It is unclear to what extent these inequalities will prove helpful, however, so we need to determine a number through trial and error.

3.3 Results and Discussion

A computational experiment was performed using the same short-haul flight delay distributions described in the in the previous chapter. A scenario composed using historical data was used to study the effect of speed control measures at a single airport. In this section, we describe the scenario and associated assumptions, we present our experimental results, and we provide some analysis.

3.3.1 Experimental Description

To conduct our studies we selected data collected from Atlanta Hartsfield-Jackson Airport on May 1, 2011. The weather conditions were clear and sunny and all runways were active. The data was obtained from an ADL file in conjunction with an ASDX file, the combination of which listed flight numbers, collection time, ETA, scheduled time of arrival (STA), the origin airport, actual time of departure, aircraft position, aircraft type, runway arrival time, STAR routes and last available fix.

The airport has 4 corner posts at the northeast, northwest, southeast and southwest corners of the airport. Arriving flights commonly fly through these corner post fixes and are sent to one of 3 runways, 2 primary runways that are used full time and another runway that is partially used.

The data was tested over a 4-hour period from 1:00-5:00 EST. CTAs were assigned using slot window sizes designed to accommodate the planned airport capacity at the time of arrival. To model the problem, we developed a simulation intended to mimic the basic effects of TMA. The simulation assumes flights proceed on their trajectories with the goal of meeting their CTAs. Once a CTA is issued, flights proceed to their assigned metering fix. When the flights reach their fixes, the simulation accepts flights for vacant runway slots on a first-come-first-served basis.

A baseline run was used to evaluate the delay performance with no intervention. This trial used flight ETAs and projected them backward to get the approximate arrival time at the metering fix. The travel times between each fix and runway were modeled by fitting flight data with separate normal distributions and sampling from these distributions. Additional uncertainty was imposed to model the variability of flights in arriving at their metering fix on time. Flights were grouped into 4 pools: Airborne flights beyond 1000 nm, airborne flights within 1000 nm, grounded flights between 500 and 1000 nm and grounded flights within 500 nm. Each pool was perturbed by sampling from a different distribution to represent the variation in travel time and received a different range of permissible arrival times. Flights beyond 500 nm were allowed to take any arrival time that could be realized solely through speed control. Flights inside of 500 nm were controlled exclusively through ground delay.

The performance was evaluated on a PC with four Intel i7-4790 dual-core processor with 8 GB of memory in a 64 bit environment. The models were coded in Python 2.7 using a GUROBI solver.

3.3.2 Results and Discussion

As we have proposed several models that feature bi-criteria objectives, we needed to identify appropriate weights for our two terms. One means of identifying the levels is to vary the weights of the term and generate a Pareto frontier. This technique is commonly referred as the weighting method. One drawback to this approach is that when the method is applied to integer programs, it can often miss points along the frontier and may generate a somewhat misleading curve. In our case, however, the stakeholders have commonly agreed upon a cost ratio of 2:1 of air to ground delay. We therefore set our weights to that level during our runs. A 15-minute look-ahead window was used on all run to better account the impact an set of assignments has on the subsequent period.

To gain a better sense of how the performance varied we ran the model at two different airport capacities and varied the maximum level of permissible ground delay. We tested four different versions of the functional approximation model: a baseline model, a model which account for equity in the objective, a model that used the by-variables and a by-variable model that incorporated our valid inequality. Our models were tested in a simulation environment using 100 monte carlo trials. The results of our tests are shown in Table 3.1. The table reports average delay transfer and the longest run time over the 4-hour assignment period for each model.

Table 3.1: A summary of model performance

Model	Max Ground Delay slots	Slotsize (min)	Number of Variables	Number of Constraints	Percentage Delay Transfer mean	Percentage Delay Transfer std	Ground-Air Ratio	Run Time (s)
FA	5	0.75	481	305	43.00	2.98	5.03	1.71
FA	7	0.75	608	327	43.73	2.43	6.13	7.77
FA	10	0.75	534	314	51.73	3.25	8.17	102.91
FA	5	0.875	425	276	26.08	3.54	4.31	0.56
FA	7	0.875	476	287	22.83	3.15	5.57	0.61
FA	10	0.875	508	288	32.25	2.52	7.38	0.84
FA with VI	5	0.75	563	520	40.31	2.77	5.11	1.98
FA with VI	7	0.75	505	508	43.70	2.55	6.47	7.21
FA with VI	10	0.75	534	511	51.14	2.88	7.66	56.97
FA-by	5	0.75	584	838	36.39	2.13	6.18	5.68
FA-by	7	0.75	594	786	43.49	2.84	7.06	88.43
FA-by	10	0.75	534	724	51.77	2.76	8.07	40.97
FA equity	5	0.75	647	324	17.90	1.29	2.51	0.30
FA equity	10	0.75	708	333	17.30	1.33	2.37	0.44
FA equity	20	0.75	801	327	16.69	1.27	2.52	0.49

The table suggests that all of our models show some ability to transfer delay away from the terminal. By allowing each flight inside of 500 nm to receive up to 5 minutes of ground delay and imposing speed control at 1000 nm we can eliminate up to 50% of the delay when the capacity is set to 80 flights per hour. Moreover, the imposed ground delay only amounts to 8% of total the airborne queuing delay in the terminal without intervention, suggesting that such flights are not inordinately penalized relative to airborne flights.

One notable facet of the results is that the different versions of the functional approximation model exhibited different levels of delay transfer when tested under identical simulation conditions. This can be attributed to the coupling between solutions as the model runs. In each instance of the problem the models solves dynamically in 15-minute intervals. Since many flights can be assigned during multiple periods, the choice of the initial solution can impose some bias on the subsequent solutions. If there are multiple optimal solutions, as is the case during the first time period, the different models may draw from different sets of assignments within the feasible region. As a result, some models will perform better than others. This difference is relatively small, however, so seemingly arbitrary choices during the initial stage will not prevent the model from transferring considerable amounts of delay. An illustration of this phenomenon is shown in delay transfer curves in Figure 3.8.

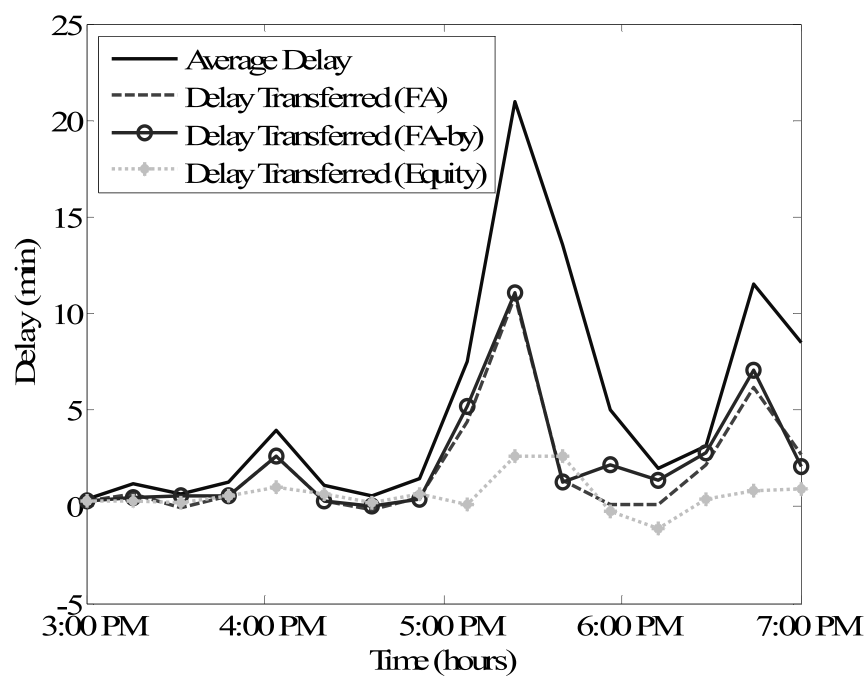

Figure 3.8 Delay Transfer Performance for each model.

One drawback to the baseline functional approximation model is that its computation time grows significantly as we impose more ground delay and a higher

AAR. While it may be impractical to impose large amount of additional ground delay on flights outside the context of a GDP or AFP it is conceivable that one might with to apply this model over multiple airports. In practice, we would only have 15 minutes to solve the problem and execute the appropriate course of action. While one could reduce the size of the look-ahead window this is not ideal as it may have a slightly negative impact on performance. Even so for this problem the solution times are not excessive. In the worst case a solution can be obtained in just over three minutes.

Since the reported run times are applied to different instances of the problem it is difficult to make an apples-to-apples comparison between models, however, we may draw certain general conclusions. The introduction of the valid inequality to the by-variable model demonstrated noticeable improvement, cutting the computation time down by roughly one half. Absent this inequality the model performs worse than the baseline model. This may be due to the strong influence the "W" variables play in influencing the solution time. The use of a scenario-based model in this context seems largely impractical due the lengthy computation time associated with the run. An effort was made to generate results, however, the results suggest that model could not be run inside of 15 minutes.

The equity-based model demonstrated stronger computational performance with decisively lower rates of delay transfer. This performance suggests that indeed the growth rate of the competing terms within the objective function do not blend well with one another. This result does not imply, however, equity could not play a factor in assigning ground delay within the system. The model could be used to generate a set of slots available to carriers within 500 nm. Once the slots were identified, a version of ration-by-schedule could be used to allocate delay to carriers.

3.4 Conclusion

In this chapter we presented three models designed to move airborne holding delay in terminal airspace to the ground and en route phase of flight. The first model was stochastic and incorporated scenarios to account for the assumed demand uncertainty. A second model used a functional approximation of the expected queuing delay that accounted for the uncertainty from multiple pools of flights. An objective which incorporated equity into the assignment of ground delay was also introduced. A third model used the Bertsimas/Stock-Patterson by-variables to assign delay an introduced a valid inequality to reduce computation time. While the first model did not meet the required computational performance the latter two models showed strong promise. Each of these two models were able to transfer up to 50% of the delay away from the terminal while not imposing significant ground delay on short-haul flights. The effectiveness was also demonstrated to a lesser extent at lower AARs.

The work in this chapter suggests that there are a number potential avenues for future study. The success in assigning a mixture of air and ground delay may prove helpful in improving GDP and AFP planning. In particular the model could be used in concert with GDPs as a measure of recourse to correct imbalances over the duration of the TMI. To do so, however, the model may need to grow some additional capability. In GDPs and AFP the airport/sector arrival rates often vary with time. A stochastic programming model could be built to handle this variation in capacity. In this context the model could be viewed as a means of dynamically adjusting the delay assigned to flights as new information evolves over the course of the TMI. From a CDM standpoint such a model could also be use facilitate additional trades between carriers. These concepts are explored in more detail in Chapter 4.

4 Delay Transfer in GDPs with Airline Control

Historically, traffic management initiatives have relied solely on ground delay to transfer delay away from the airport during inclement weather. As we have seen in the previous chapter, ground delays can effectively be used in concert with speed induced airborne delay to provide a more comprehensive balance of delay transfer to carriers. Moreover, during TMIs the need for such transfer is arguably more critical as flights participating in GDPs and AFPs often experience long delays that can produce missed connections and cancelled flights. In this chapter we studied the effect such action could have on GDPs. In section 4.1, we examine the extent to which the exemption bias between ground delayed and exempt flights can be curbed through the use of speed control. In section 4.2 we propose a more sweeping change in which we remove the exemption radius, issue CTAs to air carriers in lieu of CTDs, and allow air carriers to substitute, cancel and use speed control on the flight they operate. To study the problem, we introduce a stochastic programming model for airline decision making that allows carriers to hedge between the prospect of early weather clearance and the weather remaining in place over the duration of the GDP.

4.1 Curbing the Exemption Bias through Speed Control

Ground delay programs allow flights originating beyond a specified distance to become exempt from any delay imposed by the program. This exemption leads to a biased allocation that favors longer flights over shorter ones and alters an otherwise fair allocation. In this section we present two algorithms to reduce this exemption bias through speed control. The first algorithm attempts to assign the maximum possible delay achievable through speed control to the exempt flights. The second algorithm begins by prescribing the maximum possible delay to exempt flights, but works to improve on this allocation by acting to fill holes in the schedule with speed-controlled exempt flights whenever possible. We then present a set of experimental designs to characterize the benefit of employing such speed control algorithms to alleviate ground delay on flights.

4.1.1 Methodology

In this section we describe our methodological approach used to transfer the exemption bias to a pool of ground delayed flights. We present our two algorithms and illustrate the impact they may have on existing GDPs. The first algorithm uses two integer programming models to successively assign controlled times of arrival to exempt flights and ground delayed flights. The second algorithm simultaneously assigns arrival times to both exempt and ground delayed flights.

4.1.1.1 Delayed Exemption

There are two considerations we examined while formulating our model. As our primary goal was to aggressively transfer delay away from ground delayed flights to exempt flights, we sought to assign the maximum amount of airborne delay to exempt flights whenever possible. We also wanted to ensure the overall equitable standard was promoted by our model. While it was not possible to achieve a result as equitable as that attained in a pure RBS algorithm, we did seek to create a process in which large ground delays were discouraged whenever possible.

In order to present our model we need to define the list of sets, parameters and variables that it uses. These terms are described below.

Sets

$F \equiv$ set of all flights

Identify applicable sponsor/s here. *(sponsors)*

$E \equiv$set of all speed controlled flights

$P \equiv$set of all ground delayed flights

$S_f \equiv$ set of all slots available to flight f

$K \equiv$ set of all slots assigned to speed controlled flights

$D \equiv$ set of all slots assigned to ground delayed flights

Parameters

$t_{oagf} \equiv$The scheduled arrival time flight f

$t_s \equiv$The travel time flight of f in cruise at slot s

$t_f \equiv$The travel time flight of f in cruise at nominal fuel burn level

$\boldsymbol{\varepsilon} \equiv$A parameter between 0 and 1

Variables

$$x_{fs} = \begin{cases} 1 \text{ if flight } f \text{ is assigned to slot } s \\ \quad 0 \text{ otherwise} \end{cases}$$

IP 1:

$$\max \quad \sum_{\substack{f \in E. \\ s \in S_f}} \left(t_f - t_s\right) x_{fs} \qquad (4.1)$$

$$S.T. \quad \sum_{s \in S_f} x_{fs} = 1 \qquad \forall f \in E \qquad (4.2)$$

$$\sum_{f \in E} x_{fs} \le 1 \qquad \forall s \in S_f \qquad (4.3)$$

$$x_{fs} \in \{0,1\} \qquad (4.4)$$

Our model objective shown in equation (4.1) looks to maximize the total delay for all exempt flights. Equation (4.2) states that all exempt flights are assigned to a slot. Equation (4.3) states that no slot is occupied by more than one exempt flight. Equation (4.4) restricts our decision variables to binary values.

IP 2:

$$\min \sum_{\substack{f \in E. \\ s \in S_f}} \left(t_s - t_{oag_f}\right)^{1+\varepsilon} x_{fs} \qquad (4.5)$$

$$\text{S.T.} \quad \sum_{s \in S_f / K, s \geq t_{oagf},} x_{fs} = 1 \qquad \forall f \in P \qquad (4.6)$$

$$\sum_{f \in P} x_{fs} \leq 1 \qquad \forall s \in S_f / K \qquad (4.7)$$

$$x_{fs} \in \{0,1\} \qquad (4.8)$$

In our second IP model the objective shown in equation (4.5) aims to equitably assign delays to the pool of flights within the GDP. The coefficients of the objective are super-linear and thus they increase exponentially with large delays. This feature ensures that the model will prefer to assign small levels of delay to more flights, rather than a single large delay. Equation (4.6) states that all ground delayed flights are assigned to a slot. Equation (4.7) ensures that no slot is occupied by more than one ground-delayed flight. Equation (4.8) restricts our decision variables to binary values.

The delayed exemption algorithm uses our two integer programming models in sequential fashion. It starts by using our first model to assign delay to the exempt flights. It then uses these assignments to restrict the assignment of ground delayed flights to those that have not been assigned to the exempted flights. It then uses the second model to assign delays to flights inside the GDP radius. A description of the algorithm is shown below:

Assignment with Integer Programming Optimization

Step 1: Use IP 1 to solve for the optimal delay allocation for flights beyond the exemption radius

Step 2: Group the slots assigned to flights in IP 2 into a set of restricted slots

Step 3: Use IP 2 to solve for the optimal GDP allocation over the set of unrestricted slots

Step 4: Take $K \cup D$ and assign to flights in F based on the solutions to IPs 1 and 2

This algorithm can achieve a different allocation than the DB-RBS algorithm, as illustrated in Figure 4.1 and Figure 4.2. Under DB-RBS the two exempt flights will be assigned to the earliest available slot to their scheduled time of arrival. This results in additional delay for every flight that succeeds them even though in some cases the scheduled time of arrival for the ground delayed flights is earlier than that of the exempt flights. In the case of the delayed exemption algorithm the exempted flights receive a delay consistent with the travel at the minimum acceptable speed of the aircraft. When this delay is incurred some ground delayed flights will be allocated to slots closer to their scheduled arrival time.

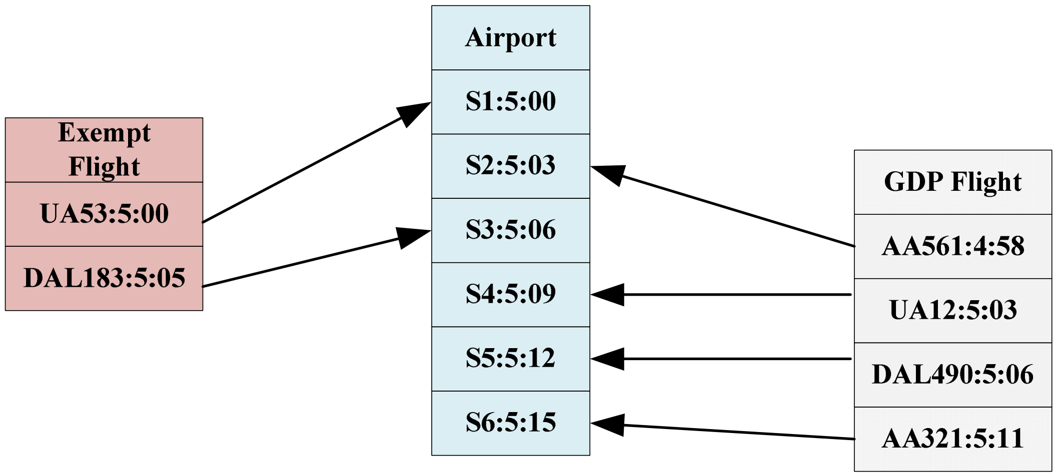

Figure 4.1 An example of flight allocation in Distance Based Ration-by-Schedule. Exempt flights receive priority.

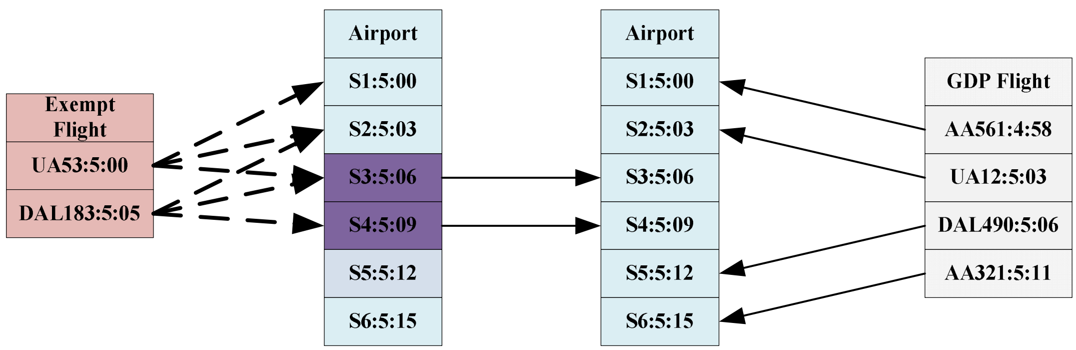

Figure 4.2 An example of flight allocation under Speed Controlled optimization procedure. Exempt flights receive the maximum possible delay and ground delay slots receive available slot based on their order in schedule.

4.1.1.2 RBS with Speed Control Exemptions

The algorithm presented in the previous section attempted to transfer delay by aggressively assigning airborne delay to exempt flights utilizing two integer programming models. While this approach certainly acts to reduce the overall level of delay for flights inside the GDP radius imposed by exemptions, it might not always yield the most efficient allocation. Moreover the transparency to stakeholders of the integer programming model it uses is questionable relative to the DB-RBS algorithm currently in place. Given these challenges we deem it worthy to present another candidate algorithm. The Speed Controlled Exemption Ration-by-Schedule

Algorithm works by initially assigning exempt flights to the maximum feasible delay and then assigning all flights to the earliest feasible slot based on their order in the schedule. When an exempt flight is assigned to a slot earlier than the maximum delay, the initial maximal delay slot enters the set of slots available for assignment. A description of the algorithm is presented below.

<u>SCE-Ration-by-Schedule Algorithm</u>

Step 1: Order flights in F by increasing scheduled time of arrival

Step 2: Select the first flight $f \in E$ that has not been assigned to a slot

a. If all flights have been assigned, group the slots into a new set R and go to Step 3

b. Otherwise, assign the flight to its last available slot in S_f

Step 3: Select the first flight in F that has not been assigned to a slot in step 2

a. If all flights have been assigned, stop and go to b.

b. Select the first available $f \in E$ move f from its previous slot s to the first available slot $a \in S_f$. If all flights have been assigned stop and exit.

c. Otherwise, assign the flight to the earliest possible unassigned slot

In addition to transferring delay this algorithm has another significant feature. Since the slot assignments of the exempted flights are not explicitly tied to the maximum possible airborne delay achievable through speed control, the model can in some instances be used to improve throughput within the system. When the call rate of the GDP is sufficiently high, gaps can sometimes emerge in which no flight can be assigned to a slot using the DB-RBS algorithm. Since DB-RBS does not assign flights to slots ahead of their scheduled time of arrival it can sometimes create situations in which some slots go unused and others are heavily desired, creating additional delays within the system. This could potentially be avoided through speed controlled flights by assigning flights ahead of their scheduled time of arrival when they are traveling at speeds below their practical limit. An example of this situation is shown in Figure 4.3 and Figure 4.4. In Figure 4.3 the DB-RBS algorithm cannot assign flights to slots that occur before their scheduled time of arrival and since there are gaps in the schedule the first slot goes unused and more delay is imposed on later flights. In the Figure 4.4 this situation is rectified by assigning exempt flights to earlier arrival times and the additional delay is avoided.

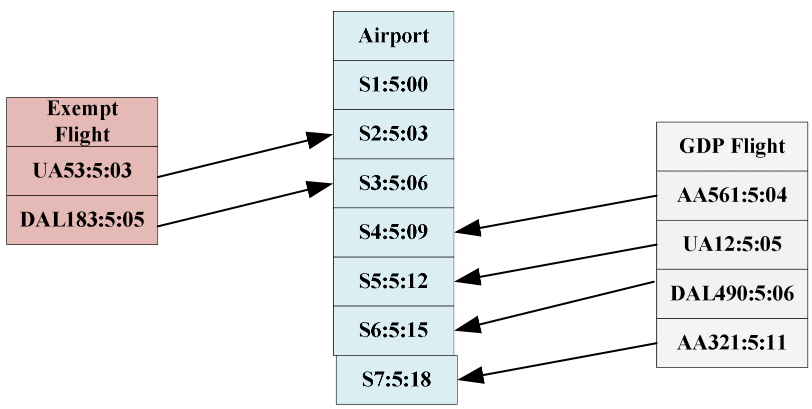

Figure 4.3 Potential slot assignment using DB-RBS under a high call rate GDP.

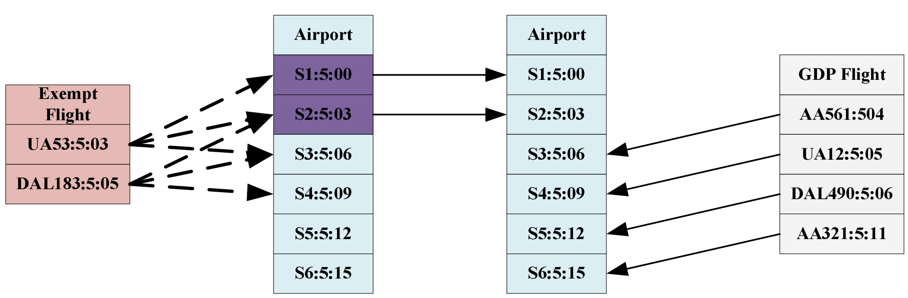

Figure 4.4 Potential slot assignment using SCE-RBS under a high call rate GDP.

4.1.2 Results and Discussion

In this section we describe a computational experiment designed to test our algorithms. The delay transfer is categorized by GDP call rate and GDP radius at three different levels. Recognizing that the compliance rate to any modification will play a crucial role in its perceived success we also examine its effect on performance.

4.1.2.1 Scenario Description

A dataset was obtained for the day of May 1, 2011 from Atlanta Hartsfield-Jackson Airport (ATL). The dataset was created by merging data from an ADL file (obtained from the FAA's Traffic Flow Management System) and an ASDX file (surface surveillance data). The key fields included: flight number, collection time stamp, expected time of arrival (ETA), scheduled time of arrival (STA), origin airport, actual time of departure, aircraft position, aircraft altitude and aircraft type.

The airport acceptance rates on an hour-by-hour basis varied from 56 to 101 flights per hour. Since this dataset was not taken on a day that an actual GDP was issued, a hypothetical GDP was superimposed on the data. A 6 hour GDP was assigned to the airport over the hours of 15:00-21:00 GMT. Flights inside the exemption radius were assigned ground delays. Flights on the ground that originated from airports outside of the radius as well as flights in the cruise phase of flight at the start of the GDP were allocated slots over the range achievable by the aircraft. The model used flight trajectories observed in the data over the day of operations. Speed control directives were issued over the period of time that the aircraft reached an altitude of 35,000 ft. Based on these trajectories we calculated the distance traveled while the aircraft was above 35,000 ft. As a baseline case flights were given a nominal cruise speed based on the aircraft performance listed on the BADA database. This database was also used to derive a set of speeds at which each aircraft could fly. In general we used these speeds as guidelines; however, speeds on all aircraft were restricted to +/-0.02 of their performance maximum/minimum. Also, when aircraft were capable of flying above Mach 0.83 or below 0.72, aircraft speeds were restricted to a maximum of 0.83 or a minimum 0.72 respectively. CTAs for ground delayed flights could correspond to any time at or following the scheduled time of arrival of the flight.

4.1.2.2 Delayed Exemption Performance

The Delayed Exemption algorithm was tested using our historical datasets. The exemption radii assumed distances of 800, 1000 and 1200 nm. These values are consistent with typical radii observed at the airport in recent years. The call rate of the GDP was examined at values of 50, 60 and 70 flights per minute. All call rates remained consistent over the lifetime of the GDP. The exemption delay was measured as the difference between the delay achieved with DB-RBS and RBS no exemptions. The delay transferred was measured as the difference in performance between DB-RBS and the Delayed Exemption algorithm. The results of our test are shown in Figure 4.5.

The algorithm performs better as the size of the exemption radius increases regardless of the call rate. This phenomenon can be explained by the change in demographics of the exempt flights. As the radius increases, the pool of flights exempt from the GDP is more likely to include transcontinental and cross country flights that have to travel significantly longer distances. These flights will spend a longer time in the air when their speed is reduced and thus more delay can be transferred.

The best performance on a percentage basis, irrespective of exemption radius, is achieved at a call rate of 60 flights/hr. This result is attributable to the arrival rate in the dataset. If the level of traffic were lighter or heavier a different call rate may have performed better.

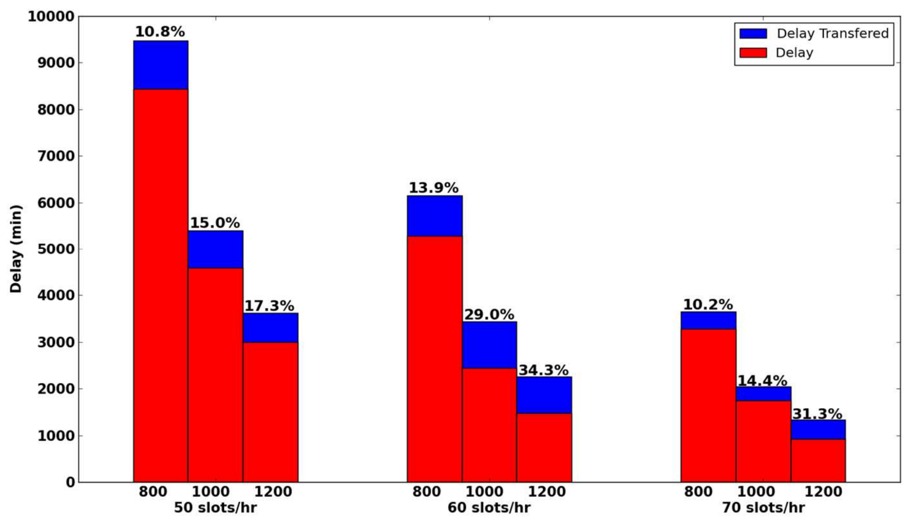

Figure 4.5 Percentage of delay transferred with Delayed Exemption algorithm at various GDP radii and call rates.

While it would be beneficial from a delay transfer standpoint to achieve near 100% compliance with speed control assignments, it is probably unrealistic to expect such a consistently high level of adherence among flights. For various reasons including making connecting flights, insufficient fuel or customer satisfaction, carriers may need to fly faster than the speed prescribed by the algorithm. As such we decided to study how the level of compliance affected performance. In this test carriers were given the option of opting out of the assigned speed controlled CTA and back into their original scheduled time of arrival. Compliance levels ranging between 0 and 100% were examined using a call rate of 60 flights/hr. The results are depicted in Figure 4.6.

Not surprisingly, the performance increases with compliance. The figure suggests that in all cases when the compliance is above 60% at least 10% of the exemption bias can be transferred. When the exemption radius is greater than 1000 nm, upwards of 20% of the delay can be transferred provided compliance remains above 80%. These results imply that a reasonable delay transfer can be achieved at suboptimal values of compliance.

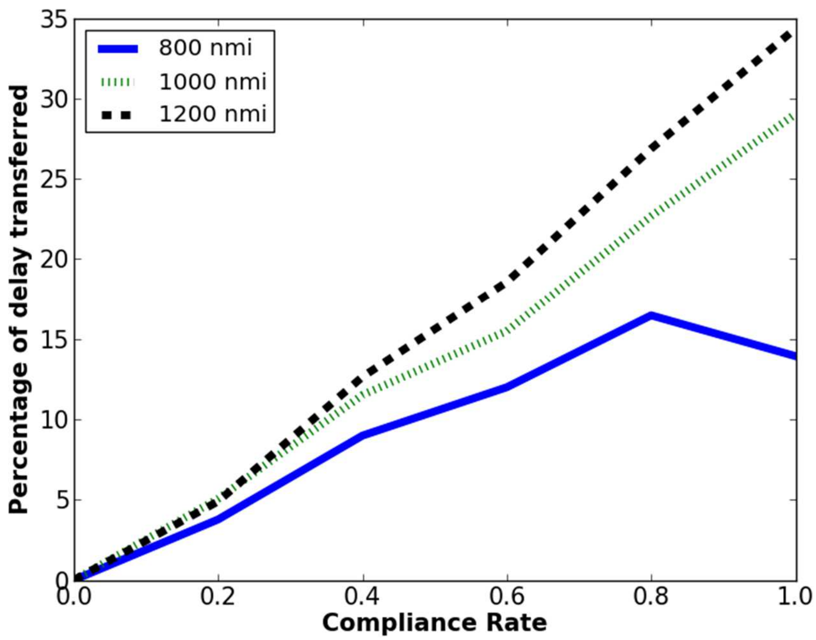

Figure 4.6 Variation in percentage of delay transferred with Delayed Exemption with carrier compliance rate.

4.1.2.3 SCE-RBS Performance

The tests performed in the previous sub-section were repeated on the SCE-RBS algorithm. The performance of the two algorithms is shown in TABLE I for comparison. The two algorithms perform comparably for call rates below 60 flights/hr. When the call rate assumes a value of 70 flights/hr the SCE-RBS algorithm outperforms the Delayed Exemption algorithm by a significant margin. This jump in performance is attributable to the fact that the algorithm is not always restricted to assigning the maximum delay to exempt flights. By occasionally speeding up flights the algorithm acts to fill gaps in the schedule and fill unused slots. In the case of the 1200 nm exemption radius the algorithm works so well it even outperforms traditional RBS. It also achieves comparable performance to the Delayed Exemption algorithm when compliance drops, as evident in Figure 4.7.

This improvement of the algorithm, however, comes with a price. In order to achieve better results some flights need to be willing to travel at faster speeds. This change may be problematic for a variety of reasons, including gate availability, as well as congestion in the terminal or en route airspace. These speed adjustments can also lead to less efficient fuel burn and while the change in small on a percentage basis it can serve as a significant cost driver. It may also be viewed as somewhat unfair to ask an airline to make the necessary accommodations to enable the flight of interest to fly at a faster speed without compensation. Mechanisms could be established, however, to reward the airline for dealing with the imposed inconvenience.

Table 4.1 Performance Exemption Delay Transferred for SCE-RBS and Delay Exemptions Algorithms.

	Delayed Exemption			SCE-RBS		
Call Rate (flights/hr)	**800 nm**	**1000 nm**	**1200 nm**	**800 nm**	**1000 nm**	**1200 nm**
50	10.8	15.0	17.3	11.0	15.0	17.4
60	13.9	29.0	34.3	21.3	28.5	33.8
70	10.2	14.4	31.23	58.3	81.4	100.5

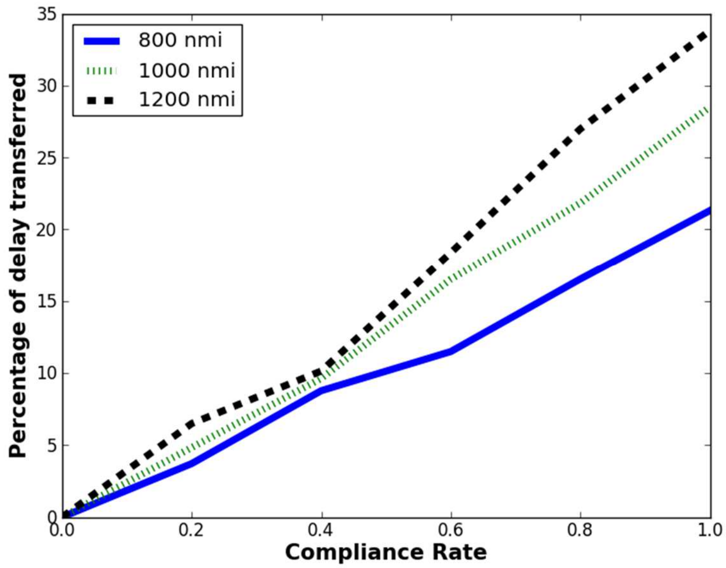

Figure 4.7 Variation in percentage of delay transferred with SCE-RBS with carrier participation rate.

4.2 Combining Speed control with CTAs in GDPs

In this chapter, we consider replacing the use of a CTD with a CTA in GDP planning and control. The principal change is conceptually quite simple: flights and, by association, flight operators, are assigned CTAs rather than CTDs. When a GDP is revised, the assigned CTAs rather than the assigned CTDs are adjusted. Because of the added flexibility provided by the use of CTAs, we also propose the elimination of GDP flight exemptions, instead allowing flight operators to effectively make exemption decisions regarding their own flights. To effect these changes, we only need to make minor changes to the existing CDM/GDP allocation procedures. We propose a new flight operator GDP planning model, specifically a scenario-based stochastic integer programming model that determines a cancellation and substitution plan for each carrier. The model matches the carrier's flights to the assigned arrival capacity (CTAs). In doing this, it takes into account the ability to adjust flight speeds en route, e.g. the model might assign a flight an "early" departure time, consistent with a relatively slow speed but anticipate the ability for the flight to increase its speed should the weather clear at the destination and additional capacity be assigned to the carrier. The integer programming models builds on the prior literature on stochastic models for GDP planning and the use of speed control extends the work of Delgado and Prats. We describe our CTA-based architecture in section 4.2.1, present our new airline optimization model to support the airline cancellation and substitution process in section 4.2.2 and describe the model used to represent compression and revisions in our experiments in section 4.2.3.

4.2.1 Architecture

The previous two sections have described in general terms the modifications that we envision to major components of the process. Here, we more specifically define the architecture and explain some of the important changes. While the new process uses the RBS mechanism, the exemption radius is eliminated. Once capacity is allocated to carriers, each carrier can use both speed control and ground delays to manage their substitution and cancellation decisions. Since no exemption radius has been imposed, carriers must be more strategic about their substitution process because in the event of an early weather clearance they will want to take advantage of capacity increases, e.g. by speeding up airborne flights. Our scenario-based stochastic model is designed to facilitate that end. Each scenario accounts for a possibility of the weather clearing at different times and the associated increase in capacity. The goal is to position the flights in slots that allow carriers to make the best use of capacity under all scenarios. The fact that slot assignment and the effective use of speed control are key to evaluating the impact of this new approach implies that proper evaluation of its effectiveness requires experiments that involve GDP revisions. The manner in which we model revisions is discussed both later in this section 4.2.2 and in section 4.2.3. Table 4.2 gives the basic steps in our CTA-based architecture.

Table 4.2 CTA-Based Flight Assignment Architecture

Step 1 [FAA].
1a: Assign a slot to each airborne flight based on the flight's expected time of arrival.
1b. Assign a slot to all flights on the ground using RBS.
1c. Create a list of slots (and CTAs) owned by each airline based on the allocation from both steps 1a and 1b.
Step 2 [Airlines]. Execute cancellation and substitution processes and adjust flight-to-CTA assignments. Assign a departure time to each flight
Step 3 [FAA]. Execute compression, adjusting assignments and filling any unusable slots.

This process looks almost identical to the existing process illustrated in Figure 1.1, however, there are some subtle differences. First, none of the flights on the ground at the start of the GDP are exempted. Second, when the airlines perform their cancellations and substitutions, and also when the FAA performs compression (steps 2 and 3), both airborne flights and flights on the ground should be considered in decision-making. The consideration of flights in the air imposes a substantial new information requirement: the (possibly very tight) limits on the degree to which their arrival times can be adjusted. Third, today, the assignment of a departure time (CTD) is performed by subtracting a nominal flight time from the CTA. Under this new approach, the airlines have substantial flexibility in assigning the departure time, e.g., as in Delgado and Prats (2012) and Delgado and Prats (2014), assuming an initial "slow" speed while anticipating possible speed-ups if weather conditions change. This added airline flexibility implies that when the airlines perform their cancellation and substitution process, they have a rich set of alternatives to consider and the opportunity to improve performance. In the next section, we present an optimization model to address this new airline decision problem.

Another very important challenge associated with this new approach is the manner in which GDP controls are dynamically updated over time. Today, a variety of possible GDP revisions might take place as weather conditions change at the destination airport. Perhaps the simplest is a cancellation of the GDP in the event of clearance of poor weather. If this occurs, all issued ground delays are immediately rescinded and the impacted flights can immediately take off. An equivalent action in a CTA-based architecture would be to allow flights on the ground to immediately depart and flights in the air increase their speed, to the extent feasible, in order to arrive at an earlier time if this is desired. It is difficult to assess a priori whether such a complete cancellation might ever be appropriate under a CTA-based system. However, it is clear that new GDP revision models and controls will be required. In particular, it is likely that "revisions" will be required not only based on major changes in conditions at the destination airport but also more minor disturbances that impact the flight times of en route flights. It is likely that such models could build on the recent experience with airborne speed control Grabbe et al (2012); Moertl (2011); Airservices Austrailia (2008); McDonald and Bronsvoort (2012); Leib (2008); Nieuwenhuisen and de Gelder (2012) and the growing body of research Knorr et al., (2011); Jones et al., (2013); Delgado and Prats (2012); Prats and Hansen (2011); Delgado and Prats (2014) on the topic. Of course, this also relates to current efforts on time-based metering and TBO.

In the current research, we have not attempted to address all the nuances of GDP revisions under CTA controls. This would certainly represent another significant research contribution. Rather, to estimate the benefits of this new architecture, we evaluate a relatively simple scenario in which weather clears at a random time and use an optimization model (Vossen & Ball, 2006a) that represents the combined effect of RBS and compression in reassigning CTAs based on the newly available capacity. This model is described in section 4.2.3.

4.2.2 Models to Support Airline Substitution and Cancellation

Under the new architecture and considering both the possibility of en route speed adjustments and no flight exemptions, each airline has more control over the disposition of its own flights. Since GDPs are often cancelled prior to their planned end time, it behooves airlines to hedge between the prospect of early and on-time cancellation. Such hedging is effectively done today by the FAA through the exemption radius. The challenge for an airline lies in positioning flights in the appropriate slots to best deal with all possibilities. To do so we adapt stochastic models developed earlier from an FAA/ANSP perspective to the perspective of a specific airline (Richetta & Odoni, 1993; Ball, et al., 2003; Ball, et al., 2010).

To understand this model, consider the deterministic case where the set of available slots, i.e. the CTAs assigned to that airline, is known with certainty, e.g. as described in (Vossen & Ball, 2006b). This is a simple assignment problem where flights are assigned to slots, allowing for the possibility that some flights may be canceled at a cost. Since this model is solved by a specific airline, we can assume the availability of a rich cost function that takes into account various factors regarding flight, crew and passenger status, passenger count, etc.

Capacity uncertainty is modeled using a set of scenarios: each scenario is characterized by the time at which that scenario becomes known, the revised set of slots, i.e. additional capacity represented by the augmentation of the existing slots with a set of additional slots, and a probability. An additional set of variables indicates how the initial assignment is adjusted when the new capacity becomes available. In defining the data underlying this model, the differences in constraints underlying airborne flights and flights on the ground must be taken into account. For example, if a flight was assigned a CTA of 4:00 and, at the time the new scenario was effective that flight was airborne, then the flight might be restricted to revised CTAs not earlier than 3:50 based on limitations on speedup options (no more than 10 minutes). On the other hand, if a flight on the ground was assigned a CTA of 4:00 and that flight still had one hour to serve on its ground delay, then that flight could be assigned any departure time within the next hour and in order to meet any new assigned CTA between 3:00 and 4:00. The air carriers should assign both a CTA and departure time to each flight. For the present experiments, we assume the departure time assigned is the earliest possible departure time that can meet the assigned CTA. This approach provides maximum flexibility where weather scenarios only allow for capacity increases. We recognize that ideally the optimization model should contain both departure time and arrival time variables – we leave such a model to future research.

This model certainly has some substantial data requirements, most notably the scenario information. There are two aspects to generating the slot lists for each scenario. The first is defining the set of *slots available to all carriers* and the second is *how those slots are assigned to each carrier*. There has been prior research on the first aspect, but this certainly would have to be adapted to this new context. For the purposes of this paper, we use representative/stylized information that captures the essential aspects of the problem setting. Regarding the second aspect, we use a basic RBS reallocation that (by necessity) cannot take into account the status (and slot assignment) of each carrier's flights. Thus, this reallocation must be viewed as an approximation; however, it only impacts cost assigned to the initial slot assignment and so it impacts only the quality of the solution and not its feasibility. We can judge the overall quality of our approach by the results of our simulation experiments. We also note that some air carriers might wish to use other processes; thus, this model could be viewed as a surrogate for any number of internal airline decision support processes.

The specific integer programming problem formulation is given below. Note that this model includes a subscript for airlines – in practice, each airline will solve its own model.

<u>Parameters:</u>

$F_a \equiv$ The set of all flights available to airline a

$A \equiv$ The set of all airlines

$S_a \equiv$ The set of all slots available to airline a

$S_{fa} \equiv$ The set of all slots available to flight f of airline a

$E_{fa} \equiv$ The set of all slots available to flight f at stage 1 prior to first probable end of the GDP available to airline a

P_{fa}≡ The set of all slots available to flight f at stage 1 following the first probable end of the GDP available to airline a

K_{fsq}≡ The set of all slots available to flight f at stage 2 in scenario q from slot s available to airline a

$d_{fsa}{}^{q}$≡ Cost of delaying flight f to slot s owned by airlines a in scenario q

c_{fa}≡The cost of cancelling flight f operated by airline a

p_q ≡ The probability of scenario q occurring

Variables:

$$x_{fsa} = \begin{cases} 1 \text{ if flight } f \text{ of airline } a \text{ is assigned to slot } s \\ \quad 0 \text{ otherwise} \end{cases}$$

$$y_{fa} = \begin{cases} 1 \text{ if flight } f \text{ of airline } a \text{ is cancelled} \\ \quad 0 \text{ otherwise} \end{cases}$$

$$z_{fsa}^{q} = \begin{cases} 1 \text{ if flight } f \text{ of airline } a \text{ is assigned to slot } s \\ \text{in scenario } q \\ \quad 0 \text{ otherwise} \end{cases}$$

$$\min \sum_{q\in Q} p_q \sum_{f\in F_a, s\in S_a} d_{fsa}^{q} z_{fsa}^{q} + \sum_{f\in F_a} c_{fa} y_{fa} \qquad (4.9)$$

$$\text{s.t.} \sum_{s\in S_{fa}} x_{fsa} + y_{fa} = 1 \quad \forall f \in F_a \qquad (4.10)$$

$$\sum_{f\in F_a} x_{fsa} \le 1 \quad \forall s \in S_{fa} \qquad (4.11)$$

$$\sum_{f\in F_a} z_{fsa}^{q} \le 1 \quad \forall s \in S_a, s \in E_{fa}, q \in Q \qquad (4.12)$$

$$x_{fsa} \le z_{fsa}^{q} \forall f \in F_a, s \in E_{fa}, q \in Q \qquad (4.13)$$

$$x_{fsa} \le \sum_{k\in K_{fsqa}} z_{fka}^{q} \quad \forall f \in F_a, s \in P_{fa}, q \in Q \qquad (4.14)$$

$$x_{fsa}, z_{fsa}, y_{fa} \in \{0,1\} \forall f \in F_a, a \in A, s \in S_{fa} \qquad (4.15)$$

Constraint (4.10) ensures that for each airline, every flight is either assigned to a slot or cancelled. Constraint (4.11) ensures that no more than one flight is assigned to a slot in the first stage of the problem. Constraint (4.12) ensures that no more than one flight is assigned to a slot in the second stage of the problem for all scenarios. Constraint (4.13) ensures that if a flight is assigned to a slot in the first stage prior to the first feasible weather clearance time, it must be assigned to the same slot in the second stage for all scenarios. Constraint (4.14) ensures that if a flight is assigned to a slot in the first stage after the first feasible weather clearance time, it must be assigned to a slot that is reachable from that slot in the second stage. Note that this constraint, through the definition of K_{fsqa}, restricts the set of slots to which a flight can be reassigned based on flight status and the various timing restrictions. Constraint (4.15) reflects that our assignment variables are binary. Our objective is to minimize the expected cost of the flight delays over all scenarios plus the cost of flight cancellations.

4.2.3 Compression and GDP Revisions

To carry out our experiments, we must both execute compression as part of the initial allocation process (see Table 4.2) and also perform a slot reallocation for the case of a GDP revision. Under current practices, revisions are performed using a modified application of RBS that takes into account both flight status and the new set of available slots. Compression is also typically performed. Very often a combined RBS/compression process is executed called RBS^{++}. In Vossen and Ball (2006b), an optimization model is defined that provides both the functionality of compression and RBS^{++}. We use this model in our experiments for both the initial compression step and also the revision process. This model actually provides carriers with more flexibility in the application of compression. However, for our purposes here, we only wish to mimic the basic processes. Specifically, the model employs a set of "goal slots," with one such slot assigned to each flight to be assigned. To mimic compression, the goal slot assigned to each flight is the RBS slot for that flight. Other assignments can be used by carriers to implement various flight prioritization schemes.

Parameters:

$F \equiv$ The set of all flights

$A \equiv$ The set of all airlines

$S \equiv$ The set of all slots

$T \equiv$ The set of all time periods

$F_a \equiv$ The set of all flights belonging to airline a

$I_a \equiv$ The set of all goal slots belonging to airline a

$R_f \equiv$ The set of slots within acceptable for flight f

$S_{ft} \equiv$ The set of all slots available to flight f in period t

$t_s \equiv$ The time corresponding to slot s

$\tau_f \equiv$ The time corresponding to the goal slot of flight f

Variables

$$x_{fsat} = \begin{cases} 1 \text{ if flight } f \text{ of airline } a \text{ is assigned to slot } s \\ \text{in time period } t \\ \qquad\qquad 0 \text{ otherwise} \end{cases}$$

$$\min \sum_{\substack{f\in F_a, s\in S_a,\\ a\in A, t\in T}} \left(\tau_f - t_s\right)^{1+\in} x_{fsat} \tag{8}$$

$$\text{s.t.} \sum_{s\in S_a, t\in T} x_{fsat} = 1 \quad \forall f \in F_a, \forall a \in A \tag{9}$$

$$\sum_{\substack{f\in F_a, a\in A,\\ t\in T}} x_{fsat} \le 1 \quad \forall s \in S \tag{10}$$

$$x_{fsat} \in \{0,1\} \quad \forall f \in F_a, \forall a \in A, \forall s \in S, \forall t \in T \tag{11}$$

Equation (9) ensures that every flight is assigned to exactly one slot. Constraint (10) ensures that each slot is assigned to no more than one flight. Constraint (11) reflects that our assignment variables are binary. The objective of the model is to lexicographically minimize the distance of the flights from their goal slots. It accomplishes this minimization by using coefficients that grow super-linearly. This mimics the impact of compression, which seeks to find a slot as close as possible to the flight's RBS slot in the case where that flight cannot be feasibly assigned to its RBS slot.

4.2.4 Results and Discussion

4.2.4.1 Experimental Description

To conduct our studies we selected data collected from Atlanta Hartsfield-Jackson Airport on May 1, 2011. The weather conditions were clear and sunny and all runways were active. The data were obtained from a file generated by TFMS in conjunction with an ASDX file, the combination of which listed flight numbers, carrier, collection time, ETA, scheduled time of arrival (STA), the origin airport, actual time of departure, aircraft position, aircraft type, arrival time.

The airport acceptance rates on an hour-by-hour basis varied from 56 to 101 flights per hour. Since this dataset was not taken on a day on which a GDP was issued, a hypothetical GDP was superimposed on the data. A 5-hour GDP was assigned to the airport over the hours of 16:00-21:00 GMT. Flights inside the exemption radius were assigned ground delays. Flights on the ground that originated from airports outside of the radius as well as flights in the cruise phase of flight at the start of the GDP were allocated slots over the range achievable by the aircraft. The model used flight trajectories observed in the data during the day of operations. Speed control directives were issued during the period of time that the aircraft reached an altitude of 35,000 ft. Based on these trajectories, we calculated the distance traveled. As a baseline, case flights were given a nominal cruise speed based on the aircraft performance listed on the BADA database. This database was also used to derive a set of speeds at which each aircraft could fly. In general, we used these speeds as guidelines; however, speeds on all aircraft were restricted to +/-0.02 of their performance maximum/minimum. Also, when aircraft were capable of flying above Mach 0.85 or below 0.72, aircraft speeds were restricted to a maximum of 0.85 or a minimum of 0.72, respectively. CTAs for ground-delayed flights could correspond to any time at or following the scheduled time of arrival of the flight.

A baseline run was used to evaluate the delay performance with no intervention. On these runs, capacity was allocated to airlines using DB-RBS. A deterministic version of the substitution and cancellation model was used; it did not account for the possibility of early clearance. A compression model was then adopted to improve throughput. To understand the full extent of the performance, we tested the computation run time of each model using a dual-core system with four Intel Xeon X5535 processors and 12 GB of memory in a 64 bit environment. The models were coded in Python 2.7 using a GUROBI solver.

4.2.4.2 The Cost of Delay and Cancellations

If this proposed scheme were implemented, each airline would compute the cost of delay based on their internal cost measures; however, to perform a computational experiment we needed to find a suitable proxy. In this paper we chose to start with the cost model presented in Pourtaklo and Ball (2009), which draws from ATA data and models from Metron Aviation. The model assumes that the direct operating cost per minute of block time is free during the first 15 minutes. After 15 minutes the cost jumps to \$64 in the air and \$32 on the ground. Since our airborne delay is essentially free from a fuel cost standpoint and fuel typically accounts for roughly half the delay costs we decided to use an equal cost for ground and air delay. Updating for yearly changes in delay costs we found the cost on both the ground and the air was \$40 (America, 2015). The Pourtaklo and Ball approach also assumes that the cost per minute of passenger delay is \$34.88 per hour or \$34.88/60 = \$0.5813 per minute. Since the airlines do not suffer the same degree of impact as customers on a per minute basis the approach approximates the cost by multiplying passenger cost by 1/6 and uses a cost of \$0.1 per minute. Adopting the same process using 2013 passenger costs we find that the additional airline cost is \$0.125 per minute, per passenger. An expression for the cost function is show below:

$$C(x,P)=\begin{cases} 0 & x<15 & (12) \\ (40+0.125P)(x-15) & 15\geq x\geq M_p & (13) \\ (40+0.125P)(M_p-15) & x>M_p & (14) \end{cases}$$

where P is the number of passengers on the flight and M_p is the maximum amount of time before the delay cost levels off. When the cost levels off it does not matter whether the airline delays the flight an additional minute or a day. Thus we assume the cost to cancel a flight is the cost at level off. Aircraft specifications were used to determine the number of passengers on a given aircraft. Using the 2013 average reported in IATA our analysis assumed a load factor 0.8 on all flights (IATA, 2014).

4.2.4.3 Cumulative Effect to Airline Cost

To evaluate the effect of the proposed GDP changes on airline costs we evaluated our expected costs in 6 different cancellation scenarios. Since seasonal variations in weather can significantly affect the probability of early GDP cancellation, we conducted a separate run with different cancellation probabilities for winter/fall, spring and summer. We assumed a cancellation threshold of 2 hours of delay. The seasonal GDP cancellation probabilities were taken from Innis and Ball (2004). A summary of our result is shown in Table 4.3. The results indicate that the potential savings ranges between 7% and 14%. The potential benefit during the spring season is more favorable due to an increased probability of early cancellation.

Table 4.3: Percentage Seasonal Cost Savings to Airlines

Cancellation Hour	**Winter/Fall**	**Spring**	**Summer**
0	19.56	22.32	22.32
1	19.46	22.21	22.21
2	17.47	19.87	19.09
3	5.51	6.25	5.72
4	1.75	1.72	1.23
>=5	1.57	1.54	1.06
Expected Cost	7.94	13.07	10.84

4.2.4.4 Effect on Airlines with no Cancellation

To better study the effect of our revision on individual airlines, we reduced the number of flights to just the 5 largest carriers. A baseline run was performed using a conventional GDP procedure. Capacity was allocated with DB-RBS and cancellations and substitutions were made using a deterministic model. Our cost function was also revised by setting M_p to a value of 90. The resulting performance for a GDP with a Planned Airport Arrival Rate (PAAR) of 40 is shown in Table 4.4. The percentage of cancellations remained relatively consistent across carriers, ranging between 25% and 33.33%. Delta and AirTran, however, exhibit stronger delay performance in traditional GDPs. This is understandable as Delta and AirTran both control a larger pool of exempt flights than regional carriers and those with a smaller presence at the airport.

To evaluate CTA-based architecture and new planning modes, we used 5 capacity profiles. The set consisted of a complete GDP and weather clearances of 15, 30, 45 and 60 minutes early. Each scenario was assumed to be equally probable. The results of our test are shown in Table 4.5. The tests yielded noticeably different results relative to the baseline. All carriers reduced their number of cancellations except for American Airlines, which only controlled 3 flights.

Table 4.4 Airline Performance with a Conventional GDP model.

Airline	**Percentage of Flights Cancelled**	**Passenger Delay**	**Number of Flights**
Delta (DAL)	27.78	11.14	108
AirTran (TRS)	32.14	12.37	28
American Southest Airlines (ASQ)	27.59	18.29	58
American (AAL)	33.33	46.50	3
Pinnacle (FLG)	25.00	29.25	4

Table 4.5 Airline Performance with a CTA-Based Architecture.

Airline	**Percentage of Flights Cancelled**	**Passenger Delay**	**Number of Flights**
Delta (DAL)	24.07	25.40	108
AirTran (TRS)	28.57	18.02	28
American Southest Airlines (ASQ)	17.24	19.24	58
American (AAL)	33.33	25.00	3
Pinnacle (FLG)	0	37.75	4

The performance data suggest that airlines will approach the two GDP procedures in remarkably different fashions. In the current framework carriers are more likely to cancel flights to create additional capacity and flexibility as well as reduce delay. In our modification, carriers have more opportunity for intra-airline substitution both through speed control and the lack of an exemption radius. The scheme also provided additional benefits in the event of an early cancellation. This is not something that is assumed in the deterministic planning case. Thus carriers will choose to keep a greater portion of their slots. Since there are far fewer cancellations, the carriers are less affected by actions of other carriers during compression. This allows carriers to have more direct control over their performance.

While the example above reveals some information regarding the relative effect of our CDM modification, it does not provide us with a sense of how strong the possibility of early clearance needs to be to affect the decision. We ran the model with another set of scenarios in which the early clearance intervals were 7.5 minutes apiece. The resulting performance of both models is shown in Figure 4.8 and Figure 4.9. In nearly all cases, the prospect of an early clearance reduced the number of cancellations while increasing the passenger delay for carriers with more long-haul flights. The magnitude of the reduction is not quite as prominent, however, as that of the 15 min scenarios.

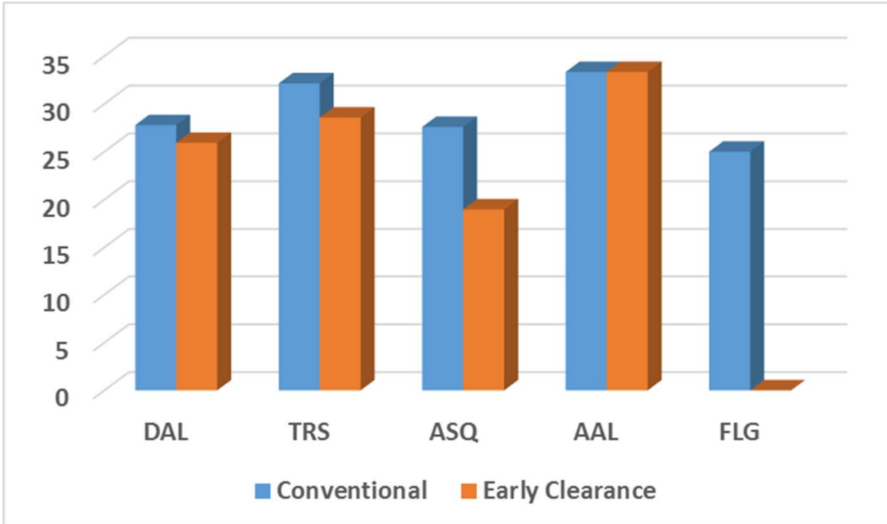

Figure 4.8 Percentage Flight cancellation level of Airlines with Conventional and Early Clearance.

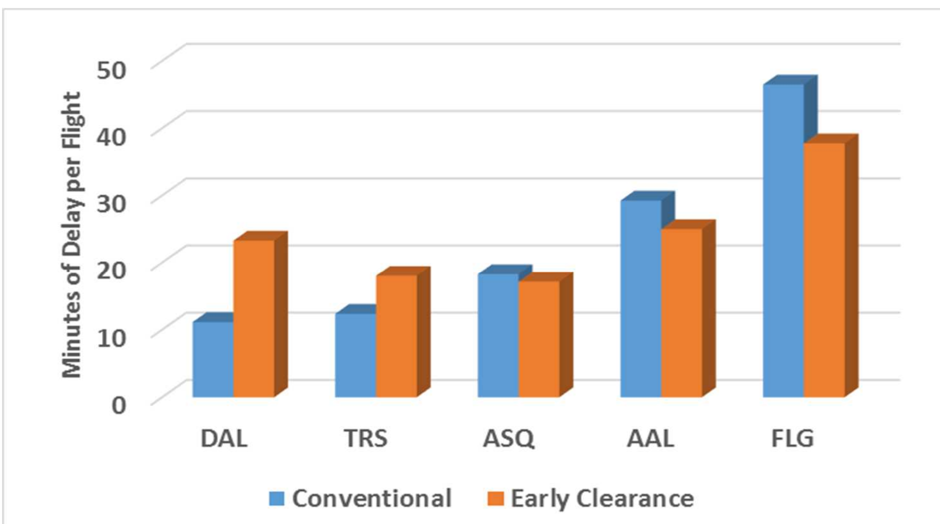

Figure 4.9 Passenger Delay of Airlines for Conventional and Early Clearance Models in minutes.

While the previous graphs demonstrate modified behavior on the part of airlines, it is unclear what portion of the change is attributable to the possibility of early cancellation vs. the lack of an exemption radius. To isolate the effect we tested our models both with and without a radius. In the former case, RBS was used to generate capacity while the later used the DB-RBS algorithm. The performance is shown in Figure 4.10 and Figure 4.11. The results suggest that when a radius is present large carriers such as Delta will reduce the number of cancellations they impose on their flights; this is also the case with Air Tran. This is likely attributable to the larger number of exempt eligible flights they have relative to other carriers. Regional carriers such as American Southeast Airlines are negatively affected by the presence of the radius and are more likely to cancel more flights to create substitution opportunities.

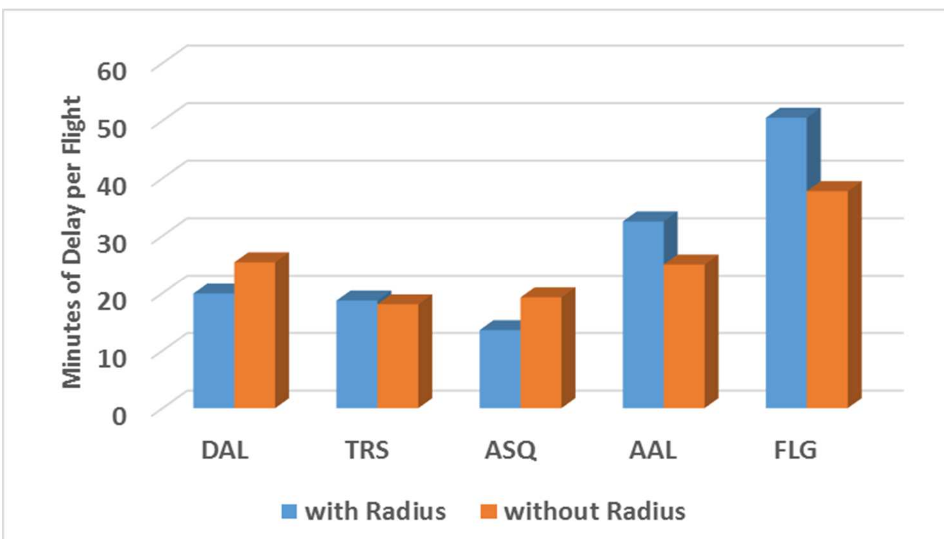

Figure 4.10 Effect of the Exemption Radius on Percentage of Flight Cancellations.

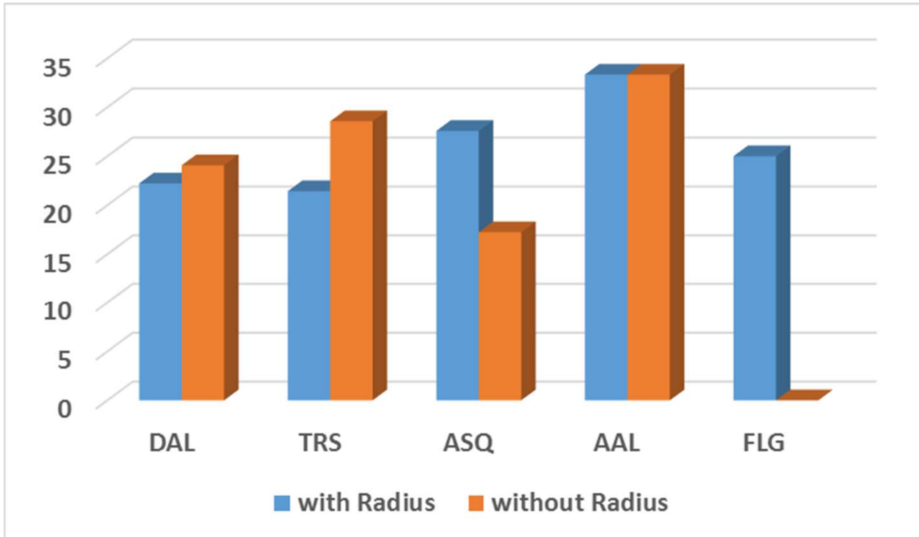

Figure 4.11 Effect of the Exemption Radius on Passenger Delay

4.2.4.5 Delay Recovery with no GDP Cancellation

We also wanted to study the potential benefit our changes could achieve in delay recovery in the event of an early GDP cancellation. To test our model we used 5 scenarios in which we assumed early clearance times of 0, 15, 30, 45 minutes and 1 hour. The performance in each case is shown in Figure 4.12 below. Delta and AirTran both experience a noticeable reduction in the overall delay as early cancellation reaches one hour. This is not entirely surprising in the case of Delta because they have a greater number of cross-continental and international flights are in a better position to recover the delay in the event of cancellation. While one might argue that carriers such as Delta operate more flights and are therefore more likely to experience more minutes of recovery, when these results are normalized on a per flight basis as shown in Figure 4.13, the benefits to long-haul carriers are still present. Thus we can conclude that our proposed revisions are beneficial to regional carriers when there is no early cancellation and beneficial to dominant and major carriers when the GDP is cancelled early.

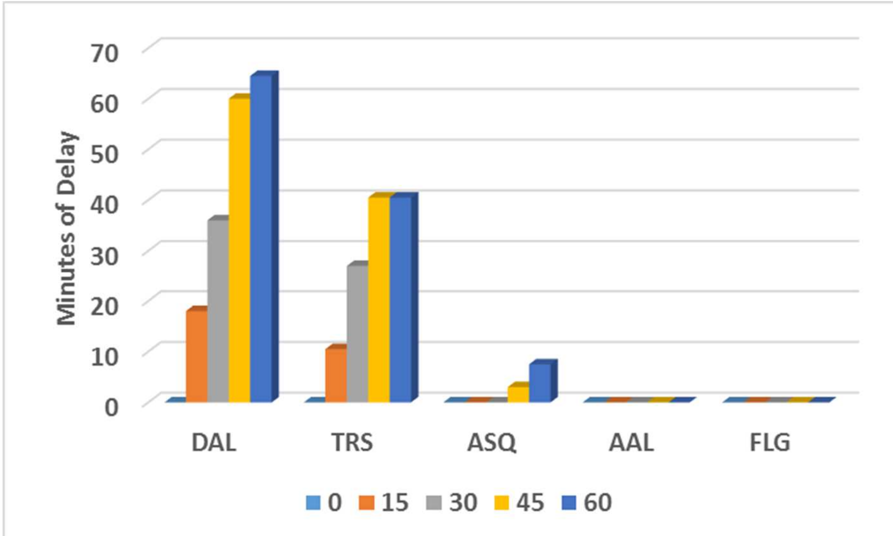

Figure 4.12 Minutes of Passenger Delay in each Scenario.

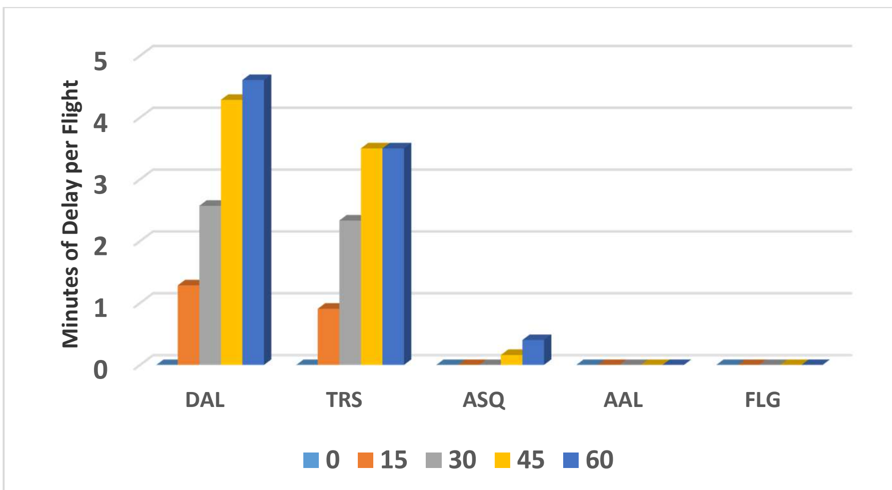

Figure 4.13: Minutes of Passenger Delay Recovered in each Scenario.

4.2.5 Conclusions and Perspectives

In this paper we proposed a new strategy for managing ground delay programs. The strategy incorporated both Controlled Departure and Arrival Times as well as en route speed control. It also eliminated the use of an exemption radius which provides incentives for carriers to create their own hedging strategies. To model performance under our new framework we adapted a stochastic model to account for airline hedging. Our analysis suggests that under our new set of GDP controls airlines are significantly less likely to cancel flights because they hope to recover delay in the event of early cancellation. Below we discuss implementation and also suggest implications on NextGen.

4.2.5.1 Near Term Implementation

The two biggest challenges to near-term implementation are i) insuring CTA compliance (as was mentioned at the outset) and ii) modifying the various GDP procedures to support the proposed architecture. Two types of enforcement can be envisioned. First, violations could be monitored and flight operators with poor records penalized in various ways. Second, as time-based metering methods are implemented CTA information could be communicated to these systems so that they could be "CTA-aware" and aid in insuring compliance. Regarding ii), the research in this chapter as well as the work on various speed control measures could be adapted to provide revision and dynamic CTA adjustment methods applicable to this context. It is probably safe to say there are no major roadblocks, just the requirement for further development and experimentation with the existing concepts. The research in this chapter also should provide a starting point for airline decision support models. A variety of approaches (some simpler, some more complex) are possible. There will be new information exchange requirements including the need for information on the limits to which CTAs can be changed for airborne flights. Of course integration with time-base metering tools would also induce new information requirements.

It should be admitted that in the near term the full benefits envisioned could not be achieved as they require complete flexibility on the part of each flight to independently adjust its speed. This limitation suggests certain NextGen goals as discussed below.

4.2.5.2 Far Term Implementation and Implications for NextGen

NextGen and Sesar both express a TBO vision in which flight timing will be closely monitored and controlled. Implicit in this vision is the ability to insure some degree of CTA compliance. In fact, one can view the architecture we have described as a (partial) vision of how GDPs would be migrated to a TBO-based NAS. NextGen and SESAR technologies also should provide the ability for flights to more independently adjust their speeds. This in turn should allow for the benefits described in this paper to be more completely realized. It is perhaps instructive to consider the underlying operational concept of our architecture. Note that, while there is a high degree of control over en route flight timing, there is also an assumption of a high degree of flexibility. This is not compatible with a TBO vision in which a 4D trajectory is set at the time of flight departure and then rigorously adhered to for the remainder of the flight. Our vision calls for a high degree of control and system-wide coordination among 4D trajectories coupled with the ability to dynamically adjust those trajectories to achieve flight operator and ANSP objectives. We feel it is important to incorporate this vision into future TBO architectures.

5 Conclusions

This dissertation made a number of contributions to further advance the use of speed control in air traffic management. First, we developed a set of strategies for transferring delay away from terminal airspace through en route speed control. To deal with the issue of uncertainty we developed a set of stochastic integer programming models that significantly increased the amount of delay transfer. These solutions were shown to work under multiple levels of system intervention. We proposed a more equitable means of allocating delay to carriers by incorporating speed control into the distance based RBS algorithm. Finally, we introduced a new architecture for GDP planning. This architecture allowed carriers to receive CTAs in lieu of CTDs and gave additional responsibilities to carriers in CDM planning. We saw that these architecture changes had the ability to noticeably reduce the expected cost of the GDP to carriers.

From an integer programming standpoint we introduced a number of models to support the application of speed control in air traffic management. The first set of models presented in chapter 2 examined transferring terminal delay when a considerable number of flights could not be controlled. To deal with the problem we proposed a model that used a value function to approximate the airborne queuing delay. The model demonstrated delay transfer levels comparable to more traditional scenario-based approaches while running two orders of magnitude faster. In chapter 3 we used a similar functional approximation model to issue a mixture of air and ground delay to a broader population of flights. The greater control of the system allowed higher levels of delay transfer with little imposition of additional ground delay. While the larger problems sizes presented some measure of stress to our computational performance, the functional approximation models were able to solve within a feasible about of time unlike the scenario-based model to which we developed as a point of comparison. In chapter 4 we adapted the use of speed control to aid in GDP planning. In section 4.1 we offered an alternative means of assigning capacity to carriers to reduce the exemption bias inherent in the current allocation method. In section 4.2 we introduced a new GDP planning procedure that allowed carriers to use both speed control and ground delay determine their preferred arrival strategies. We also proposed a new optimization model to aid carriers in determining how to most effectively reschedule their flights given an uncertain planning horizon. We then demonstrated the potential to lower the expected cost to carriers.

Through our work on these efforts we have learned a number of things regarding the effectiveness of various approaches. At the tactical level departure time uncertainty serves as a significant challenge in delivering reliable CTAs in a coordinated fashion. The presence of this uncertainty means that stochastic IP models will almost certainly significantly outperform deterministic models. While scenario-based models can prove helpful in realizing delay transfer, value function approximation models may provide comparable delay transfer at a reduced computation time. While it may be practical to develop scenarios by applying simulation tools and historical data, given the difficulty in generating large numbers of realistic scenarios, functional approximation models might offer a more practical means of reducing terminal airborne holding. Given the rich set of historical data available to practitioners, the tuning of model parameters will be instrumental in any implementation. By fitting parameter values associated with various airports and regions of airspace practitioners can gain added insight and better inform their decisions as to what situations to apply any proposed intervention.

In examining its implications on GDPs, we have also seen that speed control may be used to improve TMI operations. From the standpoint of equity, relatively simple changes to the standard allocation process could be had by modifying current allocation procedures at relatively minimal additional fuel cost to carriers. Taking such changes further and moving from CTDs to CTAs offers a more comprehensive set of system-wide benefits to carriers. These benefits include greater flexibility to carriers, a more equitable distribution of delay, fewer flight cancellations and a reduction in the expected cost to carriers. While some of these benefits will require extensions to the current time-based metering capability and an increased level of automation within the NAS, these measures could be married with TMIs as they are progressively phased in over time. Such incremental change may allow stakeholders to better adapt to the changes and implications associated with the introduction new technologies.

While we have presented some initial models aimed at addressing the use of speed control to deal with flight delay within the NAS, there are number of areas in which researchers can build upon our efforts. The functional approximation models presented in chapters 2 and 3 could be adapted to simultaneously address both demand and capacity uncertainty at the airport. In a more system-wide context, the models could also be expanded to handle the assignment of flights to multiple airports within a region. Like many other models better information at the input stage will often translate into better performance at the output stage. Along those lines, a more thorough study of departure delay characteristics at specific airports could be used to inform the level of uncertainty assumed within our models. Such information could be effectively incorporated into our model with little if any impact on our computational performance. Additionally, our study ignores the possibility of reroutes to manage delay. In practice, such reroutes are an essential component of air traffic control and offer a particularly promising area to achieve significant reductions in fuel. We eventually seek to include this capability within the model in order to provide a more comprehensive set of options to serve the needs of traffic flow managers and controllers.

There are a number of issues yet to be studied to address the need for CTAs and speed control within TMIs. Notionally the ideas we proposed to facilitate intra-airline exchange and inter-airline exchange could be paired with any number of allocation schemes including RBS, RBD and system delay oriented throughput models while injecting the use of speed control into all phases of CDM. We have assumed in our analysis that carriers and traffic flow managers and controllers conduct assignments in batches. It is quite possible and perhaps likely that stakeholders may favor a more decentralized transactional approach. Such a framework could be explored and compared as an alternative to our proposed approach. In this dissertation, we examined the use of speed control in GDPs, however, such techniques could also be used during AFPs. This added capability would provide traffic flow managers and carriers with a more rich set of options during sector-related weather events.

The notion of compliance was explored as a driver of our model performance, however, we did not take much action to more actively facilitate higher rates of compliance. One might envision a means of tracking carrier compliance and rewarding or punishing carriers that choose to heavily violate instructions by issuing them less favorable allocations in subsequent periods. Recently, flight operators and ANSPs have begun to migrate away from issuing control times of arrival toward target interval windows. In principle, one could extend our model to incorporate this idea by assigning flights to overlapping time periods. In principle these ideas could also be extended by moving away from discrete time intervals and adopting assignment using a continuous time framework. Finally, while the functional approximation model offers significantly faster solution times that scenario-based approaches the issues raised in chapter 3 suggest that if the model is to be applied to a larger setting additional steps may need to be taken to improve the computational performance. Other proxies for airborne queuing delay may aid in this endeavor.

6 Appendix A

Proposition 1: *In the single runway case the expression of expected delay shown in equations (14) and (15) serves as a lower bound of the true delay at optimality.*

Assuming we are dealing with a single period and runway, we shall define our parameters as follows:

$Q \equiv$ The set of all possible outcomes

$S \equiv$ The set of all slots

The exact expression for the expected number of arrivals in slot *s* can be written as

$$\sum_{q \in Q} p_q W_q^s = \bar{W}^s \quad (A1)$$

For a given outcome we can express the delay at slot *s* as

$$W_q^s = \max\left(\left(n_q^s + W_q^{s-1} + y_s - 1\right), 0\right) \quad (A2)$$

At optimality we have

$$W_q^{s*} = \max\left(\left(n_q^s + W_q^{s-1^*} + y_s{}^* - 1\right), 0\right) \quad (A3)$$

$V \equiv$ The set of all outcomes where $n_q^s + W_q^{s-1^*} + y_s{}^* - 1$ is positive

$U \equiv$ The set of all outcomes where W_q^{s*} is zero

Substituting the A3 into A1 we have

$$\bar{W}^s = \sum_{q \in Q} p_q W_q^s = \sum_{q \in Q} p_q \left(\max\left(\left(n_q^s + W_q^{s-1^*} + y_s{}^* - 1\right), 0\right)\right) \quad (A4)$$

$$\bar{W}^{s*} = \sum_{q \in V} p_q n_q^s + \sum_{q \in V} p_q W_q^{s-1*} + y_s{}^* \sum_{q \in V} p_q - \sum_{q \in V} p_q(1) + \sum_{q \in U} p_q(0) \quad (A5)$$

$$\bar{W}^{s*} = \sum_{q \in V} p_q n_q^s + \sum_{q \in V} p_q W_q^{s-1*} + y_s{}^* \sum_{q \in V} p_q - \sum_{q \in V} p_q(1) \quad (A6)$$

Since $n_q^s + W_q^{s-1^*} + y_s{}^* - 1$ is non-positive for all $q \in U$,

$$\bar{W}^{s*} \geq \sum_{q \in Q} p_q n_q^s + \sum_{q \in Q} p_q W_q^{s-1*} + y_s{}^* \sum_{q \in Q} p_q - \sum_{q \in Q} p_q(1) \quad (A7)$$

$$\bar{W}^{s*} \geq \bar{n}^s + \bar{W}^{s-1*} + y_s{}^* - 1 \quad (A8)$$

where $\bar{n}^s$ is the expected number of short-haul arrivals during slot *s*. We also require that the expected number of arrivals be non-negative and

$$\bar{W}^{s*} \geq 0 \quad (A9)$$

So we have

$$\bar{W}^{s*} \geq \max\left((\bar{n}^s + \bar{W}^{s-1*} + y_s{}^* - 1), 0\right) \quad (A10)$$

Since $\bar{W}^{s*}$ will always be greater than or equal to the left and right operands the expressions in equations (14) and (15) serve as a lower bound on the true expected delay.

Corollary 1: *In the single runway case the expression of expected delay shown in equations (14) and (15) serves as an exact representation of the true delay when all at least 1 flight arrives in a time interval (slot).*

The expression for the true expected delay was

$$W_q^s = \max\left(\left(n_q^s + W_q^{s-1} + y_s - 1\right), 0\right) \quad (A2)$$

When at least 1 flight arrives in every time interval (slot) the left operand is always greater than or equal to the right operand and the expression can be rewritten as

$$W_q^s = n_q^s + W_q^{s-1} + y_s - 1 \quad (A11)$$

Linearity of (A11) allows the following expectations to be taken:

$$\bar{W}^s = \bar{n}^s + \bar{W}^{s-1} + y_s - 1 \quad (A12)$$

where $\bar{n}_q^s$ is the expected number of short-haul arrivals during slot *s*. Thus $\bar{W}^s$ is an exact transformation of the true expected delay.

7 Bibliography

Agustín, A., Alonso-Ayuso, A., Escudero, L. & Pizarro, C., 2012a. On air traffic management with rerouting. Part I: Deterministic Case. *Eur. J. Oper. Res.,* 219(1), pp. 156-166.

Agustín, A., Alonso-Ayuso, A., Escudero, L. & Pizarro, C., 2012b. On air traffic flow management with rerouting. Part II: Stochastic case. *Eur. J. Oper. Res.,* 219(1), pp. 167-177.

Airbus, 2004. *Getting to grips with fuel economy, AIRBUS Flight Operations Support & Line,* Toulouse, France: Airbus.

Airlines for America, 2015. *Airlines for America.* [Online]
Available at: http://airlines.org/data/per-minute-cost-of-delays-to-u-s-airlines/
[Accessed 22 May 2015].

Airservices Austrailia, 2008. *Annual Report: 2007-2008,* s.l.: Airservices Austrailia Techinal Report.

America, A. f., 2015. *Airlines for America.* [Online]
Available at: http://airlines.org/data/per-minute-cost-of-delays-to-u-s-airlines/
[Accessed 24 January 2015].

Balakrishnan, H. & Chandran, B., 2010. Algorithms for schedule runway operations under constrained position shifting. *Operations Research,* 58(6), pp. 1650-1665.

Ball, M.; Barnhart, C.; Dresner, M.; Hansen, M.; Neels, K.; Odoni, A.; Peterson, E.; Sherry, L.; Trani, A., Zou, B., 2010. *Total delay impact study: a comprehensive assessment of the costs and impacts,* Washington, D.C: NEXTOR Report prepared for the Federal Aviation.

Ball, M. et al., 2007. Air transportation: Irregular operations and control. In: C. Barhart & G. Leporte, eds. *Handbooks in operations research and management science.* Amsterdam, NL: Elsevier B.V., pp. 1-67.

Ball, M. O., Glover, C. & Lovell, D. J., 2011. *Collaborative Approaches to the Application of Enroute Traffic Flow Management Optimization Models.* Berlin, s.n.

Ball, M. O., Hoffman, R. & Mukherjee, A., 2010. Ground delay program planning under uncertainty based on the ration-by-distance principle. *Transporation Science,* 44(1), pp. 1-14.

Ball, M. O., Hoffman, R., Odoni, A. R. & Rifkin, R., 2003. A stochastic integer program with dual network structure and its application to the ground-holding problem. *Operations Research,* 51(1), pp. 167-171.

Ball, M., Vossen, T. & Hoffman, R., 2001. *Analysis of Demand Uncertainty in Ground Delay Programs.* Santa Fe, NM, The 4th USA/Europe Air Traffic Management R&D Seminar.

Barnhart, C., Bertsimas, D., Caramanis, C. & Fearing, D., 2012. Equitable and efficient coordination in traffic flow management. *Transporation Science,* 46(2), pp. 262-280.

Baxley, B. T. et al., 2013. *Air Traffic Management Technology Demonstration-1 Concept of Operations (ATD-1 ConOps), Version 2.0.* Moffett Field, CA: NASA.

Beasley, J., Krishnamoorthy, M., Sharaiha, Y. & Abramson, D., 2000. Scheduling Aircraft Landings - The Static Case. *Transportation Science,* 34(2), pp. 180-197.

Bertsimas, D., Lulli, G. & Odoni, A., 2011. An integer optimization approach to large-scale air traffic flow management. *Operations Research,* 59(1), pp. 211-227.
Bertsimas, D. & Patterson, S. S., 1998. The air traffic flow management problem with enroute capacities. *Operations Research,* 46(3), pp. 406-422.
Bureau of Transportation Statistics, 2015. *Bureau of Transporation Statistics. U. S. Department of Transprotation. Airline On-Time statistics.* [Online]
[Accessed 22 May 2015].
Carr, G., Erzberger, H. & Neuman, F., 1998. *Airline arrival prioritization in sequencing and scheduling.* Orlando, FL, The 2nd USA/Europe Air Traffic Management R&D Seminar.
Churchill, A. M., 2010. *Coordinated and robust aviation network resource allocation. Ph.D. dissertation,* College Park, MD: University of Maryland.
Churchill, A. M. & Lovell, D. J., 2011. Coordinated aviation network resource allocation under uncertainty. *Procedia-Social and Behavioral Sciences,,* Volume 17, pp. 572-590.
De Neufville, R. & Odoni, A., 2003. *Airport Systems. Planning, Design and Management.* s.l.:Mcgraw-Hill Companies Inc..
Dear, R., 1976. *The dynamic scheduling of aircraft in the near terminal area. Research Report R76-9,* Cambridge, MA: MIT Flight Transportation Laboratory.
Delgado, L. & Prats, X., 2012. En-route speed reduction concept for absorbing air traffic flow management delays. *Journal of Aircraft,* 49(1), pp. 214-224.
Delgado, L. & Prats, X., 2013. Effect of wind on operating-cost-based cruise speed reduction for delay absorption. *IEEE Trans. Int. Transp. Sys.,* 14(2), pp. 918-927.
Delgado, L. & Prats, X., 2014. Operating cost based cruise speed reduction for ground delay programs: Effect of scope length. *Trans. Res. C: Emerg. Tech.,* Volume 48, pp. 437-452.
Evans, A., Vaze, V. & Barnhart, C., 2014. Airline-DrivenPerformance-Based Air Traffic Management: Game Theoretic Models and Multicriteria Evaluation. *Transportation Science,* p. Articles in Advance..
Gaivoronski, A. A. & Pflug, G., 2005. Value-at-risk in portfolio optimization: properties and computational approach. *Journal of Risk,* 7(2), pp. 1-31.
Ganji, M., Lovell, D. & Ball, M. O., 2009. *Resource allocation in flow-constrained areas with stochastic termination times considering both optimistic and pessimistic reroutes.* Napa, s.n.
Glover, C. N. & Ball, M. O., 2010. *Stochastic integer programming models for ground delay programs with weather uncertainty.* Budapest, s.n.
Grabbe, S., Sridhar, B., Mukherjee, A. & Morando, A., 2012. Traffic flow management impact on fuel and dealy: An Atlanta case study. *Air Traffic Control Quarterly,* 20(3), pp. 203-224.
Gupta, S., 2012. *A tractable optimization framework for Air Traffic Flow Management addressing fairness, collaboration and stochasticity. Ph.D. Disseration,* Cambridge, MA: Massachusetts Institute of Technology.
Hoffman, R. & Ball, M. O., 2000. A comparison of formulations for the single-airport ground-holding problem with banking constraints. *Operations Research,* 48(4), pp. 578-590.

IATA, 2014. *IATA.* [Online]
Available at: http://www.iata.org/pressroom/pr/Pages/2014-02-06-01.aspx
[Accessed 24 January 2015].

Inniss, T. R. & Ball, M., 2004. Estimating one-parameter airport arrival capacity distributions for air traffic flow management. *Air Traffic Control Quarterly,* 12(1), p. 223–251.

Joint Planning and Development Office, 2011. *Concept of Operations for the Next Generation Air Transportation System version 3.2,* s.l.: s.n.

Jones, J. C., Lovell, D. J. & Ball., M. O., 2013. *En route speed control methods for transfering terminal delay.* Chicago, s.n.

Knorr, D; Chen, X; Rose, M; Gulding, J; Enaud, P; Hegendoerfer, H, 2011. *Estimating ATM efficiency pools in the descent phase of flight. 9th USA/Europe Air Traffic Management Research and Development Seminar.* Berlin, Germany, s.n.

Leib, J., 2008. Flights flow get innovative fix. *Denver Post*, 26 May.

Liu, Y. & Hansen, M., 2013. Evaluation of the Performance of Ground Delay Programs. *Transportation Research Record: Journal of the Transportation Research Board,* 2400(1), pp. 54-64.

Lowther, M., Clarke, J.-P. & Ren, L., 2008. *En route speed change optimization to implement continuous descent arrivals. Proceeding of the AIAA, Guidance, Navigation and Control Conference.* Honolulu, HI, AIAA-2008-7404.

Lulli, G. & Odoni, A., 2007. The European Air Traffic Management Problem. *Transportation Science,* 41(4), pp. 431-443.

McClain, E., 2013. *Metroplex identification, evaluation, and optimization. Ph.D. dissertation,* Atlanta, GA: Georgia Institute of Technology.

McDonald, G. & Bronsvoort, J., 2012. *Concept of operations for air traffic management by managing uncertainty through multiple metering points. Air Transport and Operations: Proceedings of the Third International Air Transport and Operations Symposium.* Delft, NL, Delft University.

Miller, M. E., Capozzi, B., Guzhva, V. & Letts, B., 2014. *Airborne execution of flow strategies simulation.* Syracuse, NY, In Digital Avionics Systems Conference (DASC), 2014 IEEE/AIAA 33rd (pp. 1D1-1). IEEE..

Moertl, P. M., 2011. *Airline based en route sequencing and spacing field test results: Observations and lessons learned for extended metering.* Berlin, s.n.

Mukherjee, A. & Hansen, M., 2007. A dynamic stochastic model for the single airport ground holding problem. *Transportation Science,* 41(4), pp. 444-456.

Neuman, F. & Erzberger, H., 1991. *Analysis of delay reducing and fuel saving sequencing and spacing algorithms for arrival traffic,* s.l.: NASA Technical Memoranda.

Nieuwenhuisen, D. & de Gelder, N., 2012. *Optimizing nightly Schiphol traffic through time based operations. Air Transport and Operations: Proceedings of the Third International Air Transport and Operations Symposium.* Delft, NL, Delft Univeristy.

Odoni, A., 1987. The Flow Management Problem in Air Traffic Management. In: *Flow Control of Congested Networks.* s.l.:Springer-Verlag, pp. 269-288.

Odoni, A., 2009. Airports. In: P. Belobaba, A. Odoni & C. Barnhart, eds. *The Global Airline Industry.* West Sussex, U.K.: John Wiley & Sons Ltd., p. 363.

Pourtaklo, N. V. & Ball, M., 2009. *Equitable allocation of eroute airspace resources.* Napa, CA, The Eigth USA/Europe Air Traffic Management Research and Developement Seminar.
Prats, X. & Hansen, M., 2011. *Green delay programs, absorbing ATFM delay by flying at minimum fuel speed. 9th USA/Europe Air Traffic Management Research and Development Seminar.* Berlin, Germany, s.n.
Richetta, O. & Odoni, A., 1993. Solving optimally the static ground-holding policy problem in air traffic control. *Transportation Science,* 27(3), pp. 228-238.
Rockafellar, R. T. & Uryasev, S., 2000. Optimization of conditional value-at-risk. *Journal of Risk,* Volume 2, pp. 21-42.
Sherali, H. D., Staats, W., R. & Trani, A. A., 2003. An airspace planning and collaborative decision-making model: Part I—Probabilistic conflicts, workload, and equity considerations. *Transportation Science,* 37(4), pp. 434-456.
Sherali, H. D., Staats, W., R. & Trani, A. A., 2006. An airspace-planning and collaborative decision-making model: part II-cost model, data considerations, and computations. *Transportation Science,* 40(2), pp. 147-164.
Sherali, H. D., Staats, W., R. & Trani, A. A., 2011. Integrating slot exchange, safety, capacity, and equity mechanisms within an airspace flow program. *Transportation Science,* 45(2), pp. 271-284.
Swaroop, P. & Ball, M. O., 2013. Consensus-Building Mechanism for Setting Service Expectations in Air Traffic Flow Management. *Transportation Research Record: Journal of the Transportation Research Board,* 2325(1), pp. 87-96.
Swenson, H. et al., 2011. *Design and Evaluation of the terminal area precision and spacing system.* Berlin, s.n.
Swenson, H. et al., 1997. *Design and operational evaluation of the traffic management advisor at the Fort Worth Air Route Traffic Control Center.* Saclay, France, The 1st USA/Europe Air Traffic Management R&D Seminar.
Terrab, M. & Odoni, A. R., 1993. Strategic flow management for air traffic control. *Operations Research,* 41(1), pp. 138-152.
Tino, 2013. *Wind models and stochastic programming algorithms for en route trajectory prediction and control. Ph.D. dissertation,* Atlanta, GA: Georgia Institute of Technology.
U. S. Department of Transportation, 2015. *RITA U.S. Department of Transportation.* [Online]
Available at: http://www.rita.dot.gov/bts/press_releases/bts023_13
[Accessed 22 May 2015].
Vlachou, K., 2014. *Aviation congestion management improvements in modeling the prediction, mitigation, and evaluation of congestion in the National Airspace System. Ph.D. Disseration,* College Park, MD: Univeristy of Maryland.
Vossen, T., Ball, M., Hoffman, R. & Wambsganss, M., 2003. A general approach to equity in traffic flow management and its application to mitigating exemption bias in ground delay programs. *Air Traffic Control Quarterly,* 11(4), pp. 277-292.
Vossen, T. W. & Ball, M. O., 2006a. Optimization and mediated bartering models for Ground Delay Programs. *Naval Research Logistics,* Volume 53, pp. 75-90.
Vossen, T. W. & Ball, M. O., 2006b. Slot trading opportunities in collaborative ground delay programs. *Transporation Science,* 40(1), pp. 29-43.

Vranas, P. B., Bertsimas, D. J. & Odoni, A. R., 1994. The multi-airport ground-holding problem in air traffic control. *Operations Research,* 42(2), pp. 249-261.
Wambsganss, M., 1996. Collaborative Decision Making through dynamic information transfer. *Air Traffic Control Quarterly,* Volume 4, pp. 107-123.
Witzberger, K. et al., 2014. *NextGen technologies on the FAA's standard terminal automation replacement system.* Syracuse, NY, In Digital Avionics Systems Conference (DASC), 2014 IEEE/AIAA 33rd (pp. 1C1-1). IEEE..