The National Center of Excellence

NEXTOR

For Aviation Operations Research

Flight Predictability: Concepts, Metrics and Impacts

Final Report — February, 2014

Prepared by:

Lu Hao and Mark Hansen

University of California, Berkeley

Institute of Transportation Studies

David Lovell, Kleoniki Vlachou, and Michael Ball

University of Maryland, College Park

Institute for Systems Research

Acknowledgments

The research documented in this report was sponsored by the Federal Aviation Administration Air Traffic Organization Strategy and Performance Business Unit, through its National Center of Excellence for Aviation Operations Research (NEXTOR). The authors would like to thank David Chin and Irena Ioachim for managing and supervising this project and providing valuable feedback along the way. The authors would also like to thank Frank Ketcham, Mark Hopkins, Murray Auger, and others at Delta Airlines who contributed greatly to our understanding of predictability from the perspective of an air carrier. We also thank Megan Ryerson at the University of Pennsylvania, who was a major contributor to the fuel loading analysis. Any errors or shortcomings that remain in this report are, of course, our own.

Table of Contents

List of Figures

List of Tables

Executive Summary

In this report, we analyze two benefit mechanisms by which improved flight time reliability could reduce airline operating cost. In one mechanism, reduced variability results in short scheduled block times, and airlines realize cost savings because scheduled block time is an important driver of several categories of airline costs. In the other, reduced variability results in less contingency fuel being loaded, and airlines save on the costs to carry the extra fuel. In examining both mechanisms, we employ econometric models to determine the relationship between flight time variability and the relevant behavior (scheduled block time setting or fuel loading) and additional models to estimate the cost impacts of changes in the behavior.

We first investigate how a change in actual block time distribution will affect scheduled block time (SBT) and system performance. Firstly this relationship is studied with empirical data and multiple regression models. The distribution of the historical block time for a flight is depicted by the difference between every 10th percentile. We found that gate delay plays a minor role in setting SBT and that SBTs have decreasing sensitivity to historical flight times toward the right tail of the distribution. Using this behavior model, we generate different scenarios with hypothetical changes in actual block time distribution and evaluate the impacts of SBT, delay, and on-time performance. Focusing on the inner right tail of the distribution would yield the most improvement in the system. It is also found that even an increase in average block time can still bring benefits such as reduced SBT and better performance. Taking the heterogeneity across airlines into consideration, we also decompose the dataset to study specific airlines. Low cost carriers and United Airlines are found to be the more aggressive in setting SBT.

To measure the hypothetical economic benefits that such changes could have, we use coefficients from the open literature to relate changes in scheduled block time to reductions in scheduled flights and crew staffing. Using fleet mix distributions and cost-of ownership data, plus pilot and cabin crew salary data, we monetize those benefits. Major carriers could save on the order of $20 million with fleet usage efficiencies; similar results could not be produced for the low-cost carriers because of missing data. For crew efficiencies, however, all categories of carriers could save millions of dollars, with about three quarters of that ascribed to cockpit crew and the remainder to cabin crew.

To study fuel loading, we analyze a large and recent dataset with flight-level fuel loading and consumption information from a major US airline. With these data, firstly the relationship between the amount of loaded fuel and flight predictability performance is estimated using a statistical model. Then the impact on loaded fuel is translated into fuel consumption and ultimately, fuel cost for US domestic operations. We find that one minute of standard deviation, capturing flight unpredictability, in airborne time within a month for the same OD pair and shift of day would lead to 0.9 minute increase in loaded contingency fuel and 1.7 loaded contingency and alternate fuel.

We include a short chapter on preliminary investigations into predictability impacts of called rates for Ground Delay Programs. This effort was only gestational, so we are able to report only on some preliminary ideas for data analysis and presentation. This could be the starting point for a more exhaustive study at a later date.

Overall, we estimate annual cost savings to the US domestic airline industry on the order of $400 million would reductions in scheduled block time that could result from plausible changes in the block time distribution. We estimate fuel burn savings of $180-$240 million from eliminating flight time variability. We also find that different metrics are appropriate for these two mechanisms, with the scheduled block time based upon the inner right tail of the block time distribution (out to roughly the 70th percentile). Fuel loading considers the entire distribution, with particular focus on the outer right tail, making the standard deviation an appropriate metric.

1 Introduction

The Federal Aviation Administration (FAA), like most air navigation service providers, continuously seeks to better understand and address customer requirements, and improve the quality of service provided. Metrics for quality of the service have long been centered on delay. Thus reducing delay has been the major service quality objective. Recently, however, the concept of predictability has received more attention in service quality assessments. The FAA is seeking to define new predictability metrics so that this aspect of system performance can be monitored. The idea of predictability—also referred to as reliability or (inversely) as variability—is not a new idea in the field of ground transportation and there is extensive research in that domain on predictability concepts, measurement, and valuation. In that literature, (un)reliability mainly refers to the unpredictable variations in travel time and is thus directly related to uncertainty of travel time (Carrion and Levinson, 2012). Operationally, reliability or predictability is inversely related to dispersion of travel times between individual origin-destination pairs or on specific routes, metrics for which include variance, standard deviation, mean absolute deviation, and inter-quartile range, to name a few. Although flight predictability is still a relatively new concept, the FAA's aspiration to improve customer service through better predictability is part of its Destination 2025 plan. The Destination 2025 specific Flight Predictability performance goal is to "Improve flight predictability by reducing variances in flying time between core airports based on a 2012 baseline."(FAA, 2012).

The majority of the literature on predictability in transportation assesses predictability by measuring variability in the "travel time," which could be a road trip travel time, gate-to-gate time of a given flight, or taxi-out time of an aircraft on the airfield. There is a variety of variability measurements: difference between actual trip time and scheduled trip time (Kho et al., 2005), standard deviation of travel time distribution (Bates et al., 2001; Lomax et al., 2003; Ettema and Timmermans, 2006; Riikka and Paavilainen 2010), standard deviation over the mean travel time (Taylor 1982; Lomax et al., 2003), difference between travel time percentiles (Bolczak et al., 1997; Ettema and Timmermans, 2006; Gulding et al., 2009) and the difference in expected and actual travel delays (Cohen and Southworth, 1999; Liu and Hansen, in press). None of these studies explicitly consider the temporal aspect of predictability. In contrast, Ball et al. (2000) find that error in predicting flight departure time decreases as the departure of a flight approaches. They proposed a metric termed integrated predictive error that takes this effect into account.

Other studies are not concerned with predictability per se, but rather focus on methods for predicting travel time on the basis of the information available prior to the commencement of the travel. These studies focus on road networks and are motivated by the increasing use of routing and navigation decision support tools (Borokhov et al., 2011). Linear regression, based on a combination of the current information—system variables—and historical travel time information, has served as one of the main methodologies (Kwon et al., 2000; Zhang and Rice, 2003; Rice and Zwet, 2004). Considering the importance of prediction timeliness in the application, the algorithms are usually designed to be simple, fast and scalable (Rice and Zwet, 2004). In these studies, the travel time predictions are modeled to guide travel decisions but are not linked to performance measurement.

In the broader literature on systems, a concept closely related to predictability is entropy. Entropy has been used to measure unpredictability of a set of possible events since its introduction into information theory by Shannon (1948). Entropy has also been used to characterize stochastic processes, defined as an indexed sequence of random variables that can take values from a set of possible states. Studies have been carried out to validate the application of entropy analysis in stochastic processes (Cover and Thomas, 1991; Ciuperca and Girardin, 2005; Jacquet et al., 2008). The entropy rate is defined for all stationary processes but it is more widely linked to Markov Chain (MC) processes, in which the memoryless property leads to easy application (Cover and Thomas, 1991).

In the realm of air transportation, block time is analogous to travel time in ground transportation. Block time is the interval block timethat commences when an aircraft moves under its own power for the purpose of flight and ends when the aircraft comes to rest after landing. Block times for specific flights—e.g. United 364 from San Francisco to Washington Dulles—vary from day to day. Typically the airline scheduler sets the scheduled block time (SBT) for a certain flight more than six months ahead of time based on the estimate of the time it takes to complete each flight. Choosing the scheduled block time is similar to travelers' choice of departure time when they have a preferred arrival time. Various researchers in ground transportation have shown that travel time reliability is a significant factor that affects a traveler's departure time decision. Therefore, it is natural to assume an analogous relationship between scheduled block time and block time reliability. There are, however, few studies of how scheduled block time is decided and how the concept of predictability (reliability) is incorporated into this decision.

Scheduled block time is an important airline cost driver. Again compared to the ground transportation studies where the travel time reliability is found to have a strong effect on departure time scheduling, block time reliability is expected to be a significant factor in deciding scheduled block time. If the relationship can be understood, there may be opportunities for the FAA and other air navigation service providers to allocate resources to make block time scheduling more efficient. For example, a clearer understanding of the link between block time variability and scheduled block time might lead to the development of innovative air traffic management practices that will help improve predictability and thus allow shorter scheduled block times, while also furthering the FAA's goal of improving predictability.

The aim of this report is to contribute to the understanding of how predictability affects airline cost. In particular, we investigate two mechanisms through which improving predictability may reduce airline costs. The first is the scheduled block time effect. We conjecture that certain changes in the distribution of realized block times may reduce scheduled block times, and that the scheduled block time change, in turn, may reduce the airline operating cost. The next two chapters examine this, first by considering, in Chapter 2, the SBT effect and, in Chapter 3, the cost impact of changing SBTs. The second mechanism is related to fuel loading. Here, we hypothesize that the distribution of realized block times affects the amount of contingency fuel airlines load on flights, which in turn affects aircraft weight and fuel burn. This is the subject of Chapter 4. In Chapter 5, we present some efforts conducted as part of this project to address a different form of predictability, that of called rates for ground delay programs (GDPs). This topic is essentially unrelated to the rest of the report, although a full analysis of this subject might lead to improved benefits assessments related to day-of-operations costs incurred under traffic management initiatives (TMIs) like GDPS. Chapter 6 offers conclusions.

2 Flight Predictability and SBT Setting

2.1 Background

Block timeIn order to fully understand the processes behind block time and scheduled block time, it is instructive to review a timeline showing some of the important epochs over the span of a particular flight segment. Figure 2-1 illustrates scheduled block time (SBT) in the context of a flight time decomposition. SBT is the time duration between the scheduled (computer reservation system, or CRS) departure and scheduled arrival times. The actual block time (FT) is the time between the actual departure and arrival times and can vary from day to day for the same flight, for a variety of reasons. The block time can be further decomposed into taxi-out, airborne, and taxi-in time. The time between scheduled and actual departure time is defined as departure delay, or gate delay.

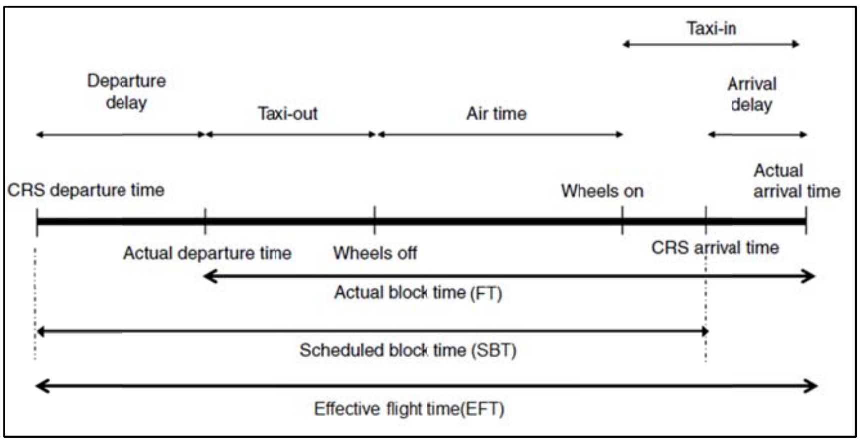

Figure 2-1: Scheduled Block Time (SBT) in the Context of Flight Time Decomposition.

SBT is an important airline cost driver. An airline's profit motive encourages a shorter SBT, because this improves fleet utilization and overall operational efficiency. However, a variety of different kinds of random events can cause flights to incur delay, which then affects on-time performance and the ability of passengers to make scheduled transfers. Because many of these delay sources are exogenous to airline operations, it is considered wise to build into the schedule some buffer time to mitigate against these delays. Therefore airlines face a difficult set of trade-offs in setting SBT. The choice of SBT is similar to a traveler's choice of departure time when they have a preferred arrival time, except the decision must be made much further in advance. Various researchers in ground transportation have shown that travel time reliability is a significant factor that affects travelers' departure time decisions. Therefore, it is natural to assume an analogous relationship between SBT and block time reliability.

In air transportation the focus of performance evaluation has long been centered on first moment metrics, such as the average actual block time. Following the spirit of the surface transportation research mentioned above, it may also be instructive to consider second moment metrics, which capture the reliability of block time. As mentioned above, a direct relationship might exist between block time reliability and SBT. To illustrate, it is observed that despite the fact that aircraft speeds have increased considerably over the last 50 years, the improvement in SBT is only marginal (citation?). One possible explanation is that increasing traffic volume is dispersing the distribution of block time, thus prolonging SBT. Nonetheless, SBT is not directly considered in most National Airspace System (NAS) cost and benefit analyses. Delay—against unimpeded block time or against the flight schedule—has been an important metric to measure the NAS performance. Delay reduction has been viewed as the major source of benefit from many initiatives to improve the NAS. However, the impact of these initiatives on SBT is not considered. Depending on this impact, a given change in block time could cause delays against schedule, and conversely the incidence of early arrivals, to increase or decrease. To be able to foresee the actual impact, we need to understand how SBT's are set and the relation between the SBT and distribution of realized block times.

In this report, the potential for changes in the block time distribution to effect a change in SBT will be the major focus. To do this, first we need to understand the relationship between SBT and actual block times. This enables our major contribution, which is to incorporate the impact of changes in the distribution of actual block times on SBT. As SBT changes, system performance metrics, such as flight delay and on-time performance will change as well. We also analyze inter-airline differences in setting SBTs and the consequent differences in performance impacts from a given change in block time distribution.

Beyond this specific focus, our study provides a perspective on how the phenomenon of transport system reliability is manifested in the specific mode of schedule air transport. As suggested above, setting SBT is somewhat analogous to scheduling the morning commute. However, there are important differences because the SBT must be set well in advance, and also in the perceived penalties of earliness and lateness. As we shall see, these differences cause airlines to focus on a particular part of the block time distribution when setting SBT. The innovative methodology required to reveal this behavior is a further contribution of our work.

2.2 Literature Review

While research into surface travel time reliability has followed many branches, we focus here on departure time scheduling when the travel time is uncertain. As mentioned above, setting SBT is roughly analogous to urban travelers choosing departure time. When travellers have a preferred arrival time (PAT), such as work start time in the morning commute, they choose their departure time based on prior knowledge of travel time on the route. The time duration between the selected departure time and preferred arrival time is fixed and serves as an implicit "scheduled" travel time. The actual travel time may deviate from this "scheduled" time from day to day due to traffic conditions, resulting in variation in actual arrival time. This analogy is further developed in Table 2-1.

In ground transportation, traveler costs due to early or late arrivals are assumed to influence departure time decisions. A traveler's PAT serves as a reference point that determines whether an arrival at a particular time is early or late. Gaver (1968) is one of the earliest advocates for this approach. In that paper, a framework for explaining variability in trip-scheduling decisions is introduced given a delay distribution and the costs of arriving early or late. Vickrey (1969) considered the tradeoff faced by travelers between queue delay and schedule delay of late or early arrivals. Knight (1974) and Pells (1987) posited the existence of a "safety margin," which is akin to a carrier scheduling a buffer against uncertain delays into the block time. Another important contribution is by Small (1982), where a theoretical model addresses departure time choice in the traveler utility function. The influence of workplace constraints on value of time is also assessed. The model proposed in Small (1982) is typically estimated using a discrete choice model based on the utility function. To expand the model of traveler choice to include uncertainty, Noland and Small (1995) expressed this uncertainty in the form of a random variable with a given probability density function. The source of the travel time dispersion (or variability) is assumed to be non-recurrent congestion. It is found that the uncertainty in travel time affects both the departure time choice and the expected costs. As uncertainty increases, travellers shift their departure time earlier to avoid late arrivals, analogous to an increase in SBT. More recent work (Fosgerau and Karlstrom, 2010) proved mathematically the statement in Bates et al. (2001), that the terms of expected earliness and lateness approximate the impact of standard deviation in the utility function.

Table 2-1: Analogy of the Travel Time Reliability Concept between Ground Transportation and Air Transportation

Concept	Ground Transportation	Air Transportation
Decision	Departure time	Block time
Scheduled travel time	Preferred arrival time minus selected departure time	Scheduled block time
Actual travel time	Actual arrival time minus selected departure time	Actual block time
Prior knowledge	Historical travel times	Historical block times
Cost of earliness/excessive scheduled block times	Lost utility from reduced time at origin	Excess labor expense, reduced aircraft utilization
Costs of lateness/insufficient block times	Late penalty, work constraints	Degraded on-time performance, traveler inconvenience, delay propagation

One attempt to predict SBT using historical data was made by Coy (2006). A two-stage statistical model of airlines' SBT is applied in the paper. Realized block time is found to be an effective predictor of SBT, having a coefficient very close to 1. In addition, arrival times, airport utilizations, and poor weather conditions were found to be significant predictors of block time. The variability (inversely reliability) of block time is not directly considered. Sohoni et al. (2011) defined two service-level metrics for an airline schedule to incorporate reliability. They also developed a stochastic integer programming formulation to adjust an existing schedule by changing departure times to maximize expected profit, while ensuring the two service levels. Chiraphadhanakul and Barnhart (2013) focus on schedule slack, defined as the additional time allocated beyond the expected time required for each aircraft connection, passenger connection, or flight leg. Considering the complexity of robust scheduling, they studied how to more effectively utilize the existing slack rather than simply having more slack to achieve a more robust schedule. Slack can absorb delay to keep the system more reliable, however at a very high cost per minute. They developed the concept of effective slack (the total aircraft/passenger slack after accounting for the historical arrival delay) with a certain upper bound, as an optimization objective.

As an effort to investigate one step beyond the first-moment metrics focusing on average block time, it is natural to consider measurements of dispersion, such as variance or standard deviation of block time, when developing a model for setting SBT. In Hao and Hansen (2013), a traditional mean-variance model is applied to capture both the centrality and dispersion of actual block time. The actual block time is decomposed into taxi-out time, airborne time and taxi-in time. Gate delay is also considered. The mean and standard deviation of each component act as explanatory variables for SBT setting. Estimation results for the dispersion term are contrary to expectations.

Capturing the unreliability of actual flight time, the standard deviations are expected to have a positive impact on SBT. However, it turns out that the standard deviations of taxi-out time and airborne time, which are the bigger sources of block time unreliability, both have negative coefficients, suggesting that an increase of the unreliability would reduce SBT for the next year, all else being equal. A one minute increase in the standard deviation would decrease SBT by 0.16 and 0.2 minutes, respectively, for taxi-out and airborne time. It is concluded from the mean-variance model that SBT is highly influenced by historical average flight times, but when these historical averages are pulled up as a result of high dispersion, the effect of dispersion is discounted. Therefore, the distribution of block time should be captured in a more detailed way than solely through second moment metrics so that the impact can be correctly identified.

As an important part in airline scheduling, the impact of SBT is profound for both airlines and the FAA. SBT directly impacts airline on-time performance. Deshpande and Arikan (2012) calculated a cost ratio of leftover (overage) cost to shortage (underage) cost, representing the relative weight airlines appear to put onto lateness and earliness of a flight. Their results show the implied flight lateness costs are less than early arrival costs for a large fraction of flights. This is different from ground transportation; however, it is consistent with Sohoni et al. (2011), which notes airlines' claim to shorten SBT in order to save cost. To do this, they are willing to incur more delay and less on-time reliability. The cost impact of SBT is shown in Zou and Hansen (2012), where econometric cost function estimates incorporating a variety of delay-buffer models reveal that both delay and schedule buffer are important cost drivers. The coefficients suggest a 0.6% increase in variable cost would occur if there were a 1 minute increase in average delay against schedule or in schedule buffer. The ability to reduce SBTs (without increasing delay against schedule) could thus result in significant cost savings.

2.3 Industry Practice

In order to capture industry practice as faithfully as possible, the research team conducted interviews with a number of major US airlines. From these interviews, we learned that historical performance data is the major source for setting SBT. Schedulers at an airline categorize the data by quarter, origin-destination pair, departure time-of-day window, and aircraft type. The time window is based on the frequency of flights and is normally 15-20 minutes. After the historical data are grouped, the primary basis for choosing SBT for a flight is the Block Time Reliability (BTR). For commercial flights, the percentile at which the SBT lies among actual block times is reported as BTR. In other words, BTR is a way to measure, for a certain flight group, how many realized flights flew a block time shorter than or equal to its SBT. BTR is different from the FAA reported on-time performance, since the on-time performance takes the whole flight, including gate delay at the origin airport, into consideration. Also, BTR does not include a 14-minute "grace period," as on-time performance does. Typically, on-time performance is not specifically considered by a carrier's block time group. Instead, it would be a flight network group's main objective to meet the on-time performance requirements. The network group works with the SBT provided by the block time group and gives feedback for SBT adjustment if they feel on-time performance will be unsatisfactory. There are intensive discussions between the two groups and the adjustment is basically reflected in the choice of the target BTR.

Typically, BTR is chosen to be in the 65th to 75th percentile of the historical block time data. Adjustments are made according to the airport characteristics, flight characteristics and feedback from other internal groups. For a major airline that has a hub-spoke network, the schedulers especially want a lower BTR for their major hub airport because there are periods of high gate utilization, and early arrivals are highly disruptive. Regarding the flight-specific characteristics, for long-haul flights, whose block time distributions tend to be more dispersed, the BTR for setting SBT is in general lower, in order to reduce average earliness. A frequent request from the network planning group is for the block time group to lower SBT, both to be more competitive with other airlines and so that there can be longer scheduled turn times. Lastly, it is worth noting that when airlines set SBT, gate delay is rarely considered in the decision process. Although gate delay clearly affects on-time performance, it is not considered part of the block time, perhaps because historical gate delay is not considered predictive of future gate delay.

2.4 Methodology and Aggregate Model

In practice, the rule for SBT setting seems to be a specific BTR (block time reliability) target. While team was informed through the interview process of the overall process for setting scheduled block times, and were also given rough ranges for some of the parameters, it was important to try to better quantify that process. Across all of the carriers we interviewed, the BTR is interpreted as a certain percentile of the historical block time distribution. Thus, we developed a model with the percentile statistics of the actual flight time. The huge amount of historical data in the field of air transportation is utilized to empirically investigate SBT setting behavior.

2.4.1 Data and Modeling

The relationship between block time distribution and block time setting is modeled empirically, using multiple regression in order to understand the relationship between SBT and past operational experience. The variables capture the difference in percentile of historical block time; therefore the model is called the percentile model in this paper. The percentile model is a generalization of the BTR target model, and assumes that, because of the adjustments to the BTR that airlines make based on on-time performance, competition, and other factors, the SBT is influenced by more than a single percentile of the historical block time distribution. For the same reason other variables than the historical distribution that might also affect the SBT decisions are also included in the model.

The data on which the SBT setting model is estimated are collected from two sources: the Airline On-time Performance dataset and the air carrier statistics data from T-100 Domestic segments with U.S. carriers, Form 41 database. Both datasets are acquired from the Bureau of Transportation Statistics (BTS). We employ the Bureau of Transportation Statistics (BTS) Airline On-time Performance data to characterize the airline schedule and operations. This database contains detailed performance information for individual flights by major US air carriers between points within the United States. These flight records are aggregated to capture the distribution of historical flight times. The aggregation of flights is by specific airlines, flight numbers, origins, and destinations: e.g. AA 112 from ORD-LGA. The time unit for the aggregation is quarter.

For each quarter, we assume that there is a uniform SBT for each individual flight, which is the elapsed time between the scheduled departure and the scheduled arrival. However, since there are occasional variations within the quarter, the median value of SBT in the quarter is used as the dependent variable. The distribution of actual flight time is captured by calculating different percentiles of the flight time data. Also, because gate delay is expected to have a different effect than flight time, we calculate the mean value of gate delay separately. For flight *f* in day *t*, the gate delay (or, departure delay) is denoted as D_{ft}^{qy}. We include $\overline{D_{f}^{qy}}$, which is the average value for D_{ft}^{qy} over the $\left|T^{qy}\right|$ days in quarter *q* of year *y*, for each flight $f \in F$, as an explanatory variable. Also, the 50th to 100th percentile of the actual flight time (block time) $Q^{f,q,y}$ is calculated. The 50th percentile or median flight time, denoted as $Q_{0.5}^{f,q,y}$, of flight $f \in F$ in quarter *q* of year *y* is included in the model. The variability of flight time is further captured by the differences between every 10th percentile from the 50th to the 100th. For example, $d_{56}^{f,q,y} = Q_{0.6}^{f,q,y} - Q_{0.5}^{f,q,y}$. This approach depicts the distribution of flight time information in a manner consistent with the industry practice mentioned above. The different segments of percentiles capture how scheduled block time is influenced by successively rarer but higher realized flight time values. To better distinguish seasonal effects, we also include dummy variables $Dummy_q^y$ for quarter q of year y.

Competition with other airlines flying the same market may motivate a shorter SBT so that the airline appears to offer faster service, or a longer SBT so that it appears more reliable. Therefore, we include variables that depict the OD pair competitiveness and airport characteristics in the model. To capture competition for the OD pair, the Herfindahl index (also known as Herfindahl–Hirschman Index, or HHI) is applied. It is an economic concept widely applied in competition law, antitrust and technology management (Liston-Hayes and Pilkington, 2004) that measures the size of firms in relation to the industry and indicates the amount of competition among them. It is defined as the sum of the squares of the market shares of the 50 largest firms (or summed over all the firms if there are fewer than 50) within the industry, where the market shares are expressed as fractions. Increases in the HHI generally indicate a decrease in competition and an increase of market power. For the purpose of our analysis, the market share of a carrier in an OD pair can be expressed as their portion of the number of seats provided out of the total number of seats serving this market. For market *od*, the HHI can be calculated as:

$$HHI_{od} = \sum_{i=1}^{N} \left(\frac{s_i}{s_{od}} \right)^2 \quad (1)$$

where s_i is the number of seats provided by carrier *i* flying this OD pair, s_{od} is the total number of seats provided in this OD pair, and N is the number of carriers serving this OD pair. Thus, in a market with two carriers where each provides 50 percent of the seats, the HHI equals $0.5^2+0.5^2 = 1/2$. A small HHI indicates a competitive industry with no dominant players. The T-100 database provides the number of seats for domestic OD pairs and carriers to calculate the HHI. Distances between origin and destination airports is also provided in the dataset and included in the model. The distance for OD pair od, in statute miles, is denoted as $dist_{od}$.

Lastly, the amount of airport traffic may also have an impact on SBT. Busier U.S. airports are more prone to congestion, and thus have larger fluctuations in block times from quarter to quarter. They also have greater gate utilization, increasing the penalty for an early arrival. In addition there are many more passengers traveling between these busier airports, and special considerations are given to SBT as they appear on the reservation system to attract customers. We therefore include dummy variables OEP_O and OEP_D indicating whether the airport is an OEP 35 airport, for origin and destination separately. The OEP 35 (Operational Evolution Partnership) airports are commercial U.S. airports with significant activity. They serve major metropolitan areas and also serve as hubs for airline operations. More than 70 percent of passengers move through these airports.

In the formulation, we assume that schedulers set SBT for a flight with the knowledge of actual flight information and the HHI competition index of the same quarter in the previous year. This implies that schedulers focus on flight experience during the same season for which they are scheduling. In this paper, data from the years 2009 and 2010 were chosen, with the SBT in 2010 modeled based on the actual flight data from the same quarter in 2009. The resulting model, with y = 2009, is:

$$SBT^{a,y+1}{}_f = \alpha_1 \times \overline{D^{qy}}_f + \alpha_2 \times dist_{od} + \beta_1 \times Q_{0.5}^{f,q,y} + \sum_{i=5}^{9} \beta_{i-3} \times d_{i,i+1}^{f,q,y} + \alpha_3 \times HHI_{od} + \sum_{q=2}^{4} \gamma_q \times Dummy^y{}_q + \gamma_5 \times OEP_O + \gamma_6 \times OEP_D + const \quad (2)$$

The flight data are filtered to include only weekday flights. To assure robustness in the data, we only include the flights that are frequently flown in a quarter. Flights flown less than 50 times on weekdays in a given quarter in either 2009 or 2010 are thus eliminated from the dataset. After this filter is applied, the estimation data set consists of 17,733 observations, each corresponding to a flight with a given flight number, operated by a given airline, between a given origin and destination.

2.4.2 Estimation Results

We estimated the model both for the entire data set and for subsets of the data corresponding to different categories of airlines. We present the results for the entire data set first. These are shown in Table 2. The coefficient of determination R^2 shows that the model explains almost 100% of the variation in scheduled block time. We can see that the coefficient for mean departure delay is quite small. A 1-minute increase of mean departure delay only suggests a change in the SBT for the next year of 0.04 minutes. Distance is positively related to SBT, indicating that airlines are being more conservative for longer flights. This suggests there is more uncertainty in longer flights that is not reflected in the historical block time distribution and SBTs are set to be longer to take the uncertainty into consideration. The impact of dispersion is captured by the coefficients of the percentile difference variables. Firstly, the impact of median flight time is 0.936, which is close to 1, indicating that this is major determinant of SBT. The $d_{i,i+1}^{f,q,y}$ variables are intended to capture the variability of flight time over the right tail of the distribution where it exceeds the median value. The interval between the 50^{th} and 60^{th} percentiles generates an increase of 0.46 minutes in SBT. The impact decreases rapidly to 0.07 minutes increase from the interval between the 70^{th} and 80^{th} percentiles and further drops to only 0.006 minutes for the right tail of the distribution. These results show that SBT is strongly affected by the left tail of the flight time distribution, the “inner right tail” has a moderate effect, whereas the additional flight time above the 70^{th} percentile has a rather small effect. This is somewhat consistent with the airline practice described in Section 2, insofar as airlines claim to choose SBT for a BTR target of around 70%. Thus, it is expected that more weight is put on the inner right tail (below the 70^{th} percentile) and the far right tail (above the 70^{th} percentile) is down-weighted. There are, however, significant differences between these results and a “pure” BTR target model, as will be discussed below.

The HHI variable has a negative coefficient. A higher HHI indicates a decrease in competitiveness for the OD pair. Thus, a negative coefficient means that if the OD market is highly competitive, airlines will increase SBT. This shows that in spite of the fact that airlines desire a shorter SBT in order to be offer faster service to their customers, they are more concerned with on-time performance.

The percentile model represents airlines’ composite SBT-setting behavior, in a manner that explicitly shows the weight they place on different regions of the historical distribution or realized block times. To further interpret the results of the percentile model, two hypothetical models for the SBT setting process are shown in the last two columns in Table 2 to compare with our estimation results.

The first hypothetical model (termed HM1) assumes that the SBT is solely determined by the average historical block time. In a CDF plot, the area above the plot corresponds to the mean value of the variable (for random variables that are strictly non-negative). Now consider a model where the mean value of realized flight time solely determines SBT. In this hypothetical model the coefficient of mean flight time would be 1. Using the CDF plot, we can translate the mean flight time into an expression based on percentile differences. If we divide the plot into the 50^{th}, 60^{th}… 100^{th} percentiles and assume the plot is piecewise linear between percentiles, then the mean value can be expressed as the sum of the areas above the CDF plot between each percentile line. For example, the area between 0 and the 50th percentile value corresponds to the contribution to the mean of the median flight time value, and can be calculated using the percentile value as the area of a trapezoid. This can be repeated for each 10 percentile interval of the tail above the 50th percentile of the distribution. The specification for hypothetical model 1 thus becomes:

$$SBT = 0.75 \times Q_{0.5} + 0.45 \times d56 + 0.35 \times d67 + 0.25 \times d78 + 0.15 \times d89 + 0.05 \times d90 \qquad (3)$$

Table 2-2: Estimation Results: Aggregate Percentile Model

	Percentile Model		**HM 1**	**HM 2**
Variable	***Estimate***	***p-Value***	***Coefficient***	***Coefficient***
Intercept	2.011	<.0001	-	-
$\overline{D_f^{qy}}$	0.039	<.0001	-	-
$\boldsymbol{dist_{od}}$	0.009	<.0001	-	-
$Q_{0.5}^{f,q,y}$	0.936	<.0001	0.75	1
$d_{56}^{f,q,y}$	0.463	<.0001	0.45	1
$d_{67}^{f,q,y}$	0.236	<.0001	0.3	1
$d_{78}^{f,q,y}$	0.075	0.0001	0.25	0
$d_{89}^{f,q,y}$	0.066	<.0001	0.15	0
$d_{90}^{f,q,y}$	0.0084	<.0001	0.05	0
$Dummy_2^y$	0.131	0.2337	-	-
$Dummy_3^y$	0.053	0.6249	-	-
$Dummy_4^y$	0.126	0.2480	-	-
$\boldsymbol{HHI_{od}}$	-2.254	<.0001	-	-
$\boldsymbol{OEP_O}$	1.037	<.0001	-	-
$\boldsymbol{OEP_D}$	0.521	<.0001	-	-
R-squared	0.9962			

Hypothetical model 2 (HM2) is a pure version of the airlines' BTR-based behavior. It assumes that SBT is equal to a certain percentile of the historical block time, for example, 70th percentile. Then the parameters of the median and the difference between 50th and 60th, 60th and 70th percentiles would be 1, since the sum of these variables is exactly the 70th percentile value, and the coefficients for the differences above 70th percentile would be 0, indicating that the airline doesn't consider the far right tail. The equation of HM2 is thus:

$$SBT = 1\times Q_{0.5} + 1\times d56 + 1\times d67 + 0\times d78 + 0\times d89 + 0\times d90 \quad (4)$$

Table 2-2 compares the results between the percentile model and the hypothetical models. HM1 only considers the mean value of flight time. In the estimated percentile model, the coefficient for the median flight time ($Q_{0.5}$) is larger in the percentile model. The coefficients for the differences from the 50th to 100th percentile decrease at a faster rate in the estimated model than they do for HM1. This clearly shows that SBTs place more weight on the left side of the flight time distribution while down-weighting the far right tail, particularly above the 70th percentile. This finding is consistent with previous literature where the implied flight delay costs are less than the implied costs of early arrivals for a large fraction of flights (Gaver, 1968). Put another way, airlines tend to be "optimistic" when they choose the SBT. They tolerate longer delays in order to realize the advantages of shorter SBTs.

HM2 assumes SBT is solely based on the 70th percentile of actual block time and thus ignores flights times beyond these values. In the estimated percentile model, the coefficient for the median value is close to 1, as in this hypothetical model. In contrast to that model, however, the inner right tail parameters are less than 1 and the outer right tail parameters are greater than 0. Thus, compared to HM2, the estimated percentile model shifts weight from the inner right to the outer right tail. One interpretation of this is that the regression model, when estimated for a large diverse set of flights, captures a composite of different BTR standards: 93% of flights have a standard at or above 50%, 46% have a standard at or above 60%, and so forth. However, it is also possible that the different regions of the block time distribution are taken into account "subconsciously" through the various adjustments airlines make to the nominal BTR standard. This seems the more likely explanation for the small but significant influence of the far right tail, since we have heard of no reports of airlines setting the BTR threshold at 80, 90, or 100%.

Returning to the comparison with the morning commute, we observe from these results that airlines are more willing to be late than most workers. While most workers would not want to be late 20% of the time, airlines pay little attention to block times over the 70th percentile. In exchange for this, they reduce earliness and avoid the high costs of setting longer block times.

2.5 Airline Specific Model

2.5.1 Model Description

Based on the study in section 2.3, we gained an overall idea about the SBT setting behavior in the industry. However, there may be heterogeneity in the SBT setting behavior across different airlines, especially between low cost carriers and legacy carriers. Legacy carriers and low cost carriers have distinct goals and strategies in their scheduling because of their different flight networks and target customers. For example, driven by the low cost goal, low cost carriers might more willingly set a shorter SBT and put less weight on the right tail of the distribution. Also, the HHI competition effect might have different impacts for low cost carriers than for legacy carriers since they view competition differently. Therefore, a study into the different airlines' behavior is conducted in this section.

In this study, we chose six U.S. carriers to represent the heterogeneity in SBT setting behavior across airlines. This includes three low cost carriers: JetBlue, Southwest, and AirTran; and three legacy carriers: American Airlines, Delta Airlines and United Airlines. The data for the three low cost airlines are aggregated because they show similar patterns, whereas three separate SBT models are estimated for the three legacy carriers.

For the legacy carriers, in addition to the independent variable that represents whether or not the origin or destination airport is large (OEP35), the hub attributes of the airport are considered in the model. Large airlines have their own hub airports where they operate a large number of flights. From our interview with airline personnel, we learned that airlines set shorter SBT for flights into their own hub airports in order to reduce early arrivals when no gate may be available, which disrupt ramp operations and annoy passengers. On the other hand, however, there are more connecting flights at the hub airports. In an effort to avoid missed connections, airlines might want longer SBT for the hub airports to assure better on-time performance. To look for these effects, we included in the legacy carrier models an additional explanatory variable indicating whether the origin or destination airport is a hub airport for the specific airline. American Airlines' hub airports include Chicago O'Hare International Airport (ORD) and Dallas/Fort Worth International Airport (DFW); Delta's hub airports include Atlanta Hartsfield-Jackson Atlanta International Airport (ATL), Minneapolis-St Paul International Airport (MSP), Detroit Metropolitan Wayne County Airport (DTW) and Salt Lake City International Airport (SLC); United's hub airports include San Francisco International Airport (SFO), Chicao O'Hare International Airport (ORD), Washington Dulles International Airport (IAD) and Denver International Airport (DEN). The variables *Hub_origin* and *Hub_des* are dummy variables that indicate this airport attribute.

2.5.2 Estimation Results

The estimation results for the airline specific model are listed in Table 2-3. For the low cost carriers and American Airlines, the flight time variables have very similar results to the overall model in section 2.4. The mean gate delay again has a very small but positive impact, and distance also has positive impact. The median value is a major driver and the impact of historical block time distribution attenuates rapidly along the right tail. However, the pattern is not quite the same for Delta Airlines and United Airlines. Mean gate delay and the intercept are not significant in the model for Delta and mean gate delay is not significant for United. Regarding the flight time distribution, median historical flight time is still a major contributor for Delta. The inner right tail (up to the 80^{th} percentile) of the historical flight time has a large and significant coefficient, whereas the percentile differences beyond 80^{th} percentile are no longer significant. For United Airlines, median value is again a major predictor, however only the percentiles up to the 60^{th} are significant in their SBT setting model. This indicates that United Airlines is unusually aggressive when it sets SBTs and gives little consideration to the right tail of the distribution. While being aware that the actual block time will often be longer then the SBT they set, United Airlines is more willing to take that risk and suffer potential delay.

Regarding the OD pair characteristic variables, whether or not the airport is large (OEP 35) seems not to matter for most of the cases. Low cost carriers is the only group which shows an impact of the OEP destination dummy on SBT, and the effect is negative. Low cost carriers set shorter SBT when their flights are flying into a large airport. This is probably a strategy small airlines are using to appear more attractive to customers on the market by having a shorter SBT in the reservation system. The competition index HHI has a significant effect on SBT for Delta and American Airlines, but not for the low cost carriers and United. Low cost carriers as well as United thus do not appear to consider the competitiveness for a certain market when setting SBT. For Delta Airlines, the coefficient is positive. This indicates that higher competition in the market would drive Delta to reduce their SBT, a point made in the airline interviews. For American Airlines, the coefficient is negative. The more competitive the market, the longer American Airlines sets its SBT. American's response suggests that it considers on-time performance to be a more effective means of attaining market share than short scheduled flight durations.

Table 2-3: Estimation Results: Airline Analysis

	Percentile Model				Hypothetical Model	
Variable	***LCC***	***AA***	***DL***	***UA***	***1***	***2***
Intercept	1.909 (3.75)	2.304 (2.35)	1.773 (1.10)	5.632 (8.66)	-	-
$\overline{D_f^{qy}}$	0.037 (4.47)	0.056 (3.50)	0.035 (1.07)	-0.014 (-1.01)	-	-
$\boldsymbol{dist_{od}}$	0.0046 (4.61)	0.0027 (2.41)	0.00496 (2.59)	0.006 (7.47)	-	-
$Q_{0.5}^{f,q,y}$	0.967 (114.3)	0.985 (102)	0.966 (58.47)	0.964 (133.63)	0.75	1
$d_{56}^{f,q,y}$	0.472 (6.84)	0.748 (7.22)	0.461 (3.02)	0.3797 (4.84)	0.45	1
$d_{67}^{f,q,y}$	0.249 (4.43)	0.271 (3.26)	0.677 (4.85)	0.068 (1.04)	0.3	1
$d_{78}^{f,q,y}$	0.147 (3.41)	0.205 (3.53)	0.256 (2.57)	0.0046 (0.09)	0.25	0
$d_{89}^{f,q,y}$	0.078 (3.34)	0.083 (2.5)	0.0674 (1.14)	-0.0455 (-1.62)	0.15	0
$d_{90}^{f,q,y}$	0.0035 (1.03)	0.027 (4.78)	-0.0108 (-1.11)	0.0024 (0.58)	0.05	0
Q_2^y	0.2438 (0.99)	-0.925 (-2.46)	-3.697 (-5.34)	3.242 (10.12)	-	-
Q_3^y	-0.6997 (-3.00)	-0.065 (-0.16)	-3.129 (-4.80)	1.262 (4.15)	-	-
Q_4^y	-1.8596 (-7.16)	-2.322 (-5.76)	-4.0397 (-5.95)	3.616 (12.34)	-	-
$\boldsymbol{HHI_{od}}$	0.9595 (2.18)	-2.187 (-3.97)	2.481 (1.80)	-0.491 (-0.88)	-	-
$\boldsymbol{OEP_O}$	0.316 (1.42)	0.365 (0.6)	-0.222 (-0.27)	0.387 (0.99)	-	-
$\boldsymbol{OEP_D}$	-0.935 (-4.34)	-0.459 (-1.16)	1.191 (1.32)	1.068 (2.75)	-	-
Hub_origin	-	-1.398 (-4.2)	-2.308 (-3.66)	-0.321 (-1.03)	-	-
Hub_des	-	-1.459 (-3.73)	-1.882 (-3.23)	-0.799 (-2.59)	-	-
R-squared	0.9967	0.9955	0.9962	0.9976	-	-
No. of observations	2363	1825	586	1978	-	-

The dummy variable indicating hub airports only applies to the legacy carriers. The coefficients are all significant (except for *Hub_origin* for United) and negative, showing that airlines want shorter SBT for flights involving their own hub airports. The *Hub-des* results are again consistent with the industry practice of seeking to avoid early arrivals at hubs in order to avoid disrupting gate operations described in section 2.3. There is no obvious explanation for the *Hub_origin* results, however.

2.6 Impact Study

2.6.1 Methodology

With the understanding of the SBT setting behavior gained above, we now consider how a hypothetical change in the block time distribution affects SBT, and hence various metrics for performance. From the percentile model, the median value and the difference between every 10th percentiles above the median captures the contribution from actual block time. The FAA and airlines are constantly seeking Air Traffic Management (ATM) procedures to improve actual block time. The improvements include surface management to reduce excessive taxi time and air traffic control to lessen airborne delay. All of these changes will result in a change in the historical block time distribution for each individual flight over a certain time period. This will further cause the SBT of the flight to change, and delay and on-time performance metrics accordingly also change. To explore this, we hypothesize changes in the block time distribution, and then assess their impacts.

In general, the scenario changes all have the same basic form. For each individual flight $f \in F$ in quarter *q*, the actual block time *T* is sorted, meaning $T_{1,f,q}$ is the smallest block time value, and $T_{n_{f,q},f,q}$ is the largest, where $n_{f,q}$ equals the total number of observations for this flight. We developed an algorithm that applies changes to the difference between consecutively increasing values of the distribution. The change is illustrated by multiplying a coefficient to the original difference as in equation 5, where T' indicates the adjusted value of block time:

$$T'_{i+1,f,q} - T'_{i,f,q} = k_{i,f,q}(T_{i+1,f,q} - T_{i,f,q}), i = 1,...,n_{f,q} \qquad (5)$$

For every flight and all the scenarios, the median block time $T_{\frac{n_{f,q}}{2},f,q}$ remains the same. In the case where *n* is odd, the two values in the middle of the distribution both remain the same. The change is then applied step-wise towards the left and right tails. The scenario is captured in the coefficient k_i, which varies systematically across the distribution. The restriction $0 < k_i < 1$ guarantees that the difference between block times will be smaller after adjustment. In other words, the adjustment is meant to push the distribution of block time towards the center (median) by reducing the difference between consecutive values, from both tails.

In scenario 1, the change in actual block time difference is a small fraction of the original difference at the far left tail. Moving toward the right, k_i increases and the change gets smaller. Specifically, in scenario 1, k_i is a linear function of *i*: $k_{i,f,q} = \frac{i}{n_{f,q}}$. Scenario 1 depicts a condition where the left tail of the historical block time distribution is more aggressively changed than the right tail. Thus scenario 1 reduces flight time variability primarily by increasing the shortest flight times. Scenario 2 is the opposite of scenario 1, with the largest change on the right tail of the distribution. Thus we set $k_{i,f,q} = 1 - \frac{i}{n_{f,q}}$. This implies that the longest block times are reduced the most. Scenario 3 assumes a consistent change throughout the distribution. In scenario 3, k_i is set to be 0.5, meaning the new differences between actual block times are all half of the original.

Figure 2 illustrates the different scenarios. The blue line indicates the CDF of the original block time for an example flight. In all three scenarios the dispersion of the original distribution is reduced, albeit in different ways. In scenario 1, illustrated in green, the left tail is pushed more, while in scenario 2 in red the right tail is shifted more. In scenario 3, the shifts are equal for both tails.

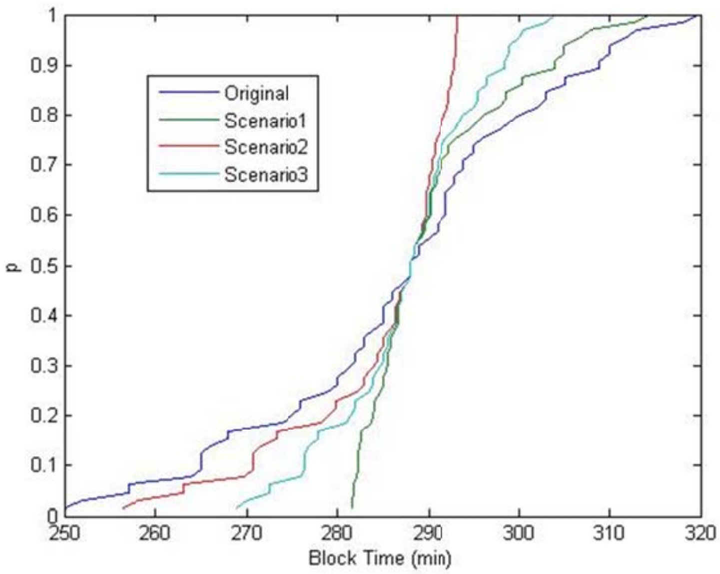

Figure 2-2: CDF Plot of Actual Block Time under Different Scenarios

Based on these scenarios block times for the 17,733 individual flights studied previously were recalculated. Based on these distributions, we recalculate the difference between every 10th percentile and acquire the data we need to recalculate SBTs with the percentile model as in equation 2. In the percentile model, only the variables related to actual block time distributions are changed. Other variables such as flight and airport characteristics, as well as gate delay, remain the same. Then, for the simulated flight with a new SBT and a new actual block time, we calculate average reduction in SBT and various other metrics, across all 17,733 flights.

2.6.2 Aggregate Model

Table 2-4 shows the results for various performance metrics under the three scenarios. To identify the effect of changed SBT specifically, we make a comparison calculation where the block time distribution is changed as above while SBT remains the same, as shown in the right half of the table. For the three scenarios, we keep the median block time the same. Thus, the change in average block time would be different for the three scenarios. The top part of Table 2-4 illustrates the difference in average block time of each scenario compared to the original data. Scenario 1 increases average block time by 1.03 minutes per flight, while scenario 2 and 3 reduce average mean block time by 3.08 and 1.02 minute per flight. These results are intuitive given the nature of the scenarios.

As for the reduction in SBT, scenario 2 has the largest reduction of 2.27 minutes per flight, followed by 1.71 minutes in scenario 3, and 1.11 minutes in scenario 1. In the percentile model for setting SBT, the left tail is only captured by the median value. Since the median value remains the same for the three scenarios, scenario 2 applies the largest reduction to the right tail, while scenario 3 applies the smallest. Therefor the results are within expectation. In scenario 1, even though the average block time is increased, the SBT still goes down since the left tail is still shifted inward somewhat. This illustrates the importance of determining how a change in the NAS will affect the distribution of block times, rather than merely its first moment, in order to assess its impacts on SBT.

Table 2-4: Evaluation Results under Different Scenarios

		With SBT change			**Without SBT change**		
	Original	*Scenario1*	*Scenario2*	*Scenario3*	*Scenario1*	*Scenario2*	*Scenario3*
Mean Block Time	144.53	145.56	141.45	143.51	145.56	141.45	143.51
Difference		1.03	-3.08	-1.02	1.03	-3.08	-1.02
SBT	146.54	145.43	144.27	144.83	146.54	146.54	146.54
Difference		-1.11	-2.27	-1.71	0	0	0
Delay	-2.03	0.13	-2.82	-1.34	-1.00	-5.11	-3.05538
Difference		2.17	-0.79	0.69	1.03	-3.08	-1.02
Positive Delay	4.22	2.98	0.71	1.83	3.24	0.92	1.98
Difference		-1.24	-3.51	-2.39	-0.98	-3.30	-2.24
Negative Delay (Absolute Value)	6.25	2.85	3.53	3.17	4.24	6.03	5.04
Difference		-3.41	-2.72	-3.08	-2.01	-0.22	-1.22
On-time performance	0.803	0.815	0.854	0.841	0.815	0.860	0.838
Improvement		0.013	0.0518	0.038	0.013	0.058	0.036
A0 on-time performance	0.570	0.593	0.661	0.617	0.592	0.681	0.625
Improvement		0.023	0.091	0.0470	0.022	0.112	0.055

Two system-wide performance metrics, delay and on-time performance, are evaluated in this study. Delay is defined as the difference between actual block time and SBT. It is further decomposed as positive delay and negative delay. Positive delay counts negative delay as zero thus captures the lateness of flights, while negative delay counts positive delay as zero and captures the earliness. Scenario 2 (1) more aggressively shortens the right (left) tail of the distribution, i.e., the longer (shorter) block times in the distribution. Therefore scenario 2(1) reduces the late (early) arriving flights most significantly, whereas the early (late) arriving flights are not reduced greatly because they are less aggressively shortened. Therefore, reflected in overall delay, scenario 2 reduces the overall delay by 0.79 minutes per flight while scenarios 1 and 3 bothincrease average delays. Scenario 2 reduces more positive delay while scenarios 1 and 3 reduce more negative delay.

Regarding on-time performance, both the A14 and A0 metrics are evaluated. A flight is on time under A14 (A0) standard if it arrives less than 15 (0) minutes after scheduled arrival time. In this study, we use the same gate delay for each realized flight. The only change is in the actual and scheduled block time. The bottom part of table 4 shows the results. On average 80.3% of flights in a quarter arrive on-time under A14 and 57% under A0. For all three scenarios, there is improvement in the on-time performance. Scenario 2 again has the largest improvement, followed by scenarios 3 and 1. This is because scenario 2 reduces the most positive delay. Although the overall delay for scenarios 1 and 3 is not decreased, the on-time performance is still better. The increase in delay is mainly because the earliness is reduced more. Therefore, in on-time performance reflecting only the lateness of flights, scenarios 1 and 3 also show improvement. However, the overall improvement in on-time performance is rather small, because another major factor that impairs on-time performance, the gate delay of the flight, is not changed under these scenarios.

Lastly, under all the scenarios, if we only apply the change to actual block time and ignore the change in SBT that this might bring, there is substantial difference in the performance metrics. For scenario 3, the delay will appear to decrease without a change in SBT, while considering SBT change the delay would actually increase. The on-time performance metric also yields a slightly more "optimistic" result without considering SBT change. The gap between these two groups indicates that the change in SBT should be taken into consideration when assessing the impact of changes in the NAS.

2.6.3 Airline Specific Model

Similar to the analysis in section 2.4, we want to take the heterogeneity across airlines into consideration as well when evaluating the impact of changed block time distribution. The data are again stratified into a low cost carrier group and three legacy carriers. In this airline specific analysis, we take the same approach as above. The change in SBT under a given scenario is computed for each airline group, leading to the change in delay and on-time performance. Since we already know that not considering SBT change leads to biased results, only the impacts with SBT change considered are presented in Table 2-5. Under the same three scenarios as above, the changes in average block time and SBT follow the same patterns as the population model, where scenario 1 brings up the average block time and scenarios 2 and 3 bring it down. All three scenarios manage to reduce SBT. American and Delta reduce SBT more than low cost carriers, while United only has a minor reduction in SBT. This reflects the results of the SBT models, showing that American and Delta give more consideration to the right tail of the block time distribution, while United sets SBT more aggressively and cares less about the right tail. Therefore, the change in actual block time would affect United less because a large part of the distribution is neglected in their SBT setting.

The delay is decomposed into positive and negative delay. All four airline groups could reduce positive delay (reflecting lateness) on a similar scale. Regarding negative delay (reflecting earliness), United yields the least reduction in negative delay among the four groups, and scenario 2 does not reduce negative delay at all for this carrier. Since United is an aggressive SBT scheduler, they will eagerly cut SBT to avoid early arrivals, thus absorbing the negative delay change that the other airlines will experience. Reflected in overall delay, low cost carriers reduce more positive delay and thus have a reduced overall delay. Overall delay for American and Delta is not reduced after the block time adjustment. The reduction in earliness is greater than the reduction in lateness so the overall delay is increased. These two legacy carriers are identified as more conservative carriers who care more about the right tail of the distribution, therefore the adjustment results in more improvement in the early arrivals rather than the late arrivals. On the contrary, United reduces more lateness than earliness, thus has a reduction in overall delay. This reduction happens under scenario 2 and 3 where the right tail of the block time distribution is reduced significantly. Lastly, for on-time performance, all four groups experience roughly the same impact. The improvement in on-time performance exists on a similar scale across airlines but is very minor compared to the change in delay.

Table 2-5: Evaluation Results: Airline Analysis

	LCC			*DL*			*AA*			*UA*		
	1	*2*	*3*	*1*	*2*	*3*	*1*	*2*	*3*	*1*	*2*	*3*
Mean Block Time Difference	0.96	-3.27	-1.16	1.19	-3.44	-1.12	1.23	-3.58	-1.17	1.31	-3.35	-1.02
SBT Difference	-1.82	-3.03	-2.38	-2.16	-3.88	-3.02	-1.93	-4.51	-3.22	-0.19	-0.27	-0.23
Delay Difference	2.74	-0.28	1.23	3.35	0.44	1.90	3.17	0.93	2.02	1.51	-3.08	-0.79
Positive Delay Difference	-0.91	-3.32	-2.13	-1.00	-3.37	-2.21	-1.17	-3.45	-2.33	-1.20	-3.02	-2.34
Negative Delay Difference	-3.65	-3.04	-3.35	-4.35	-3.81	-4.10	-4.34	-4.38	-4.35	-2.71	0.06	-1.56
On-time performance Improvement	0.01	0.05	0.03	0.01	0.06	0.04	0.01	0.05	0.03	0.02	0.05	0.04
A0 on-time performance Improvement	-0.01	0.05	0.01	0.00	0.06	0.02	0.01	0.04	0.02	0.05	0.14	0.09

2.7 Summary

In this section, we study the impact of historical block time distribution (reflecting block time reliability) on SBT, as well as system-wide performance. According to the airlines, SBT is set using a BTR target, which acts as a metric for block time reliability. We developed the percentile model in order to capture the airlines' BTR based practice. The variability in block time is captured by increments between every 10th percentile above the 50th. This enables us to observe how different regions of the historical block time distribution are considered in SBT setting. Other variables, such as gate delay, distance, airport size and hub status, and competiveness are also included in the model. To simulate change in actual block time, different scenarios are assumed that change different regions of the block time distribution differently. The change is mapped back into SBT change, and then to various delay and on-time performance metrics. The analysis is conducted on an aggregate data set as well as for individual carriers and carrier groups.

At the aggregate level, the behavior model suggests that the entire right tail of the block time distribution is considered when setting SBT, but that the inner right tail receives the most consideration. The far right tail of the historical block time distribution only has a minor impact on SBT. In general, airlines are willing to experience delay in exchange for a shorter SBT. There is substantial variation among airlines, with UA among the more aggressive and AA among the least. Thus UA appears willing to risk more delay to keep SBTs low, while AA is willing to set higher SBTs in order to increase reliability. Among other factors, notable results include that historical gate delay is virtually ignored, that airlines with hubs tend to set shorter SBTs for their hub-bound flights, and that the impact of competition varies across airlines. Delta and low cost carriers shorten the SBT when dealing with high competition, while AA chooses to lengthen their SBTs.

This model draws an explicit connection between SBT and the historical distribution of realized block times. With the model, we then studied how a change in this distribution, for example as the result of NEXTGEN improvements, will affect SBT, and therefore delay against schedule as well. It is clear from our results that knowledge of the change in *average* block times is not sufficient for this, since a given change in the average can arise from many different changes in the distribution. From our results, even an increased average block time can result in reduced SBT and improved system performance. The system performance improvement is mainly reflected in delay reduction, yet not quite significant in on-time performance. This suggests that business cases for NAS improvements should pay more attention to impacts on the distribution of block times, instead of the average. Our results indicate that focusing on the right tail, i.e. the excessive block times, would bring the best performance results. When the FAA and the airlines are considering potential procedures for block time improvement, it is necessary to take the corresponding change in SBT into consideration. Failure to do so may result in skewed results. Lastly, while it is presumptuous for external analysts to tell airlines how they should set SBTs, it is curious that historical gate delay is largely ignored in this process. A greater improvement in on-time performance might be attained by giving more consideration to this factor, to the extent it can be predicted for past experience.

3 SBT Benefits Assessment

In chapter 2, the impact of block time predictability on SBT setting is explored with econometric models. The anticipated mechanisms by which benefits could be realized as a result of improvements in strategic flight predictability can be articulated as follows:

- A reduction in the variability of actual flight times should lead to a reduction in scheduled block times and fuel buffers.
- The reduction in scheduled block times should lead to shorter actual block times.
- The reduction in fuel buffer will lead to a reduction in contingency fuel loaded, which will also lead to a reduction in actual fuel usage.
- With improvements in scheduled and actual block times, carriers could hypothetically achieve the same levels of scheduled operations with fewer aircraft and less total crew duty time.
- While the number and duration of operations is not expected to change under this hypothesis, the fuel burned on every segment of each itinerary would be reduced.

The work flow for the benefits assessment begin with the effort described above, in which the team at U.C. Berkeley developed airline specific scenarios for the impact of changes in empirical block time distributions on scheduled and actual block times. The Maryland team used these scenarios as the entry points for an economic benefits assessment. This work flow is shown in Figure 3-1 below.

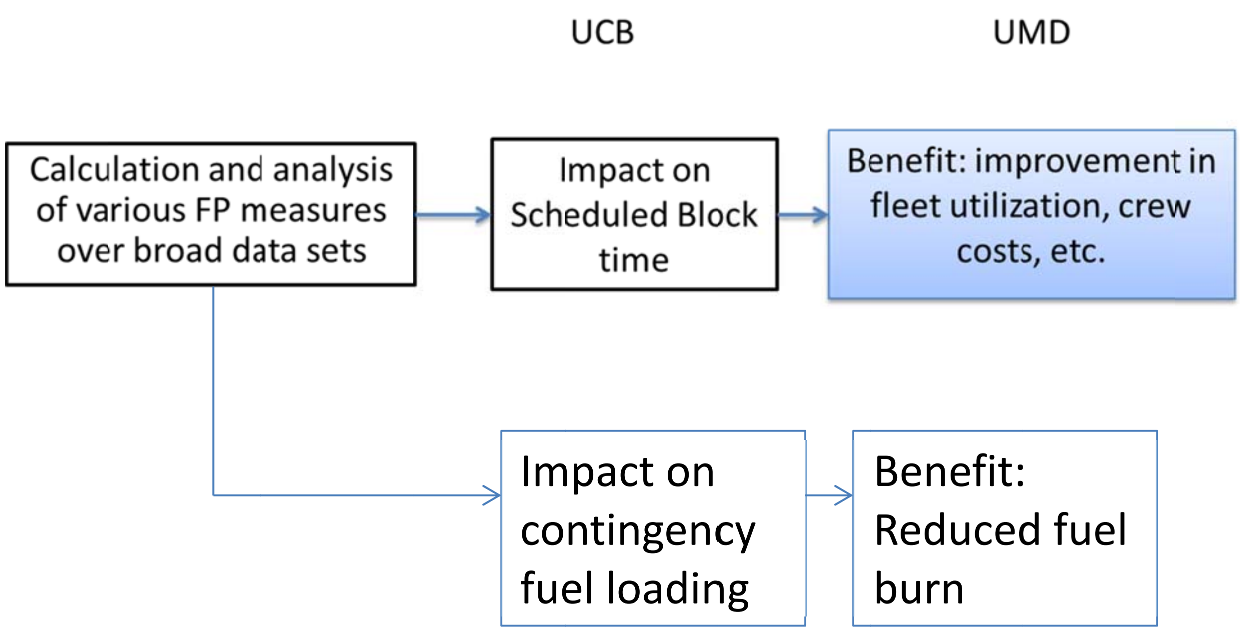

Figure 3-1: Benefits assessment work flow

The benefits assessment process used the results from the first two rows of Table 2-5 above. These changes in mean actual and scheduled block time can be related to changes in average daily flights per aircraft and average pilot salary per available seat mile (ASM) using regression coefficients from a study by Moreno-Hines and Kirkman (2013). Other data sources include publicly available data from the Bureau of Transportation Statistics (BTS). Figure 3-2 shows a flowchart of the benefits assessment methodology.

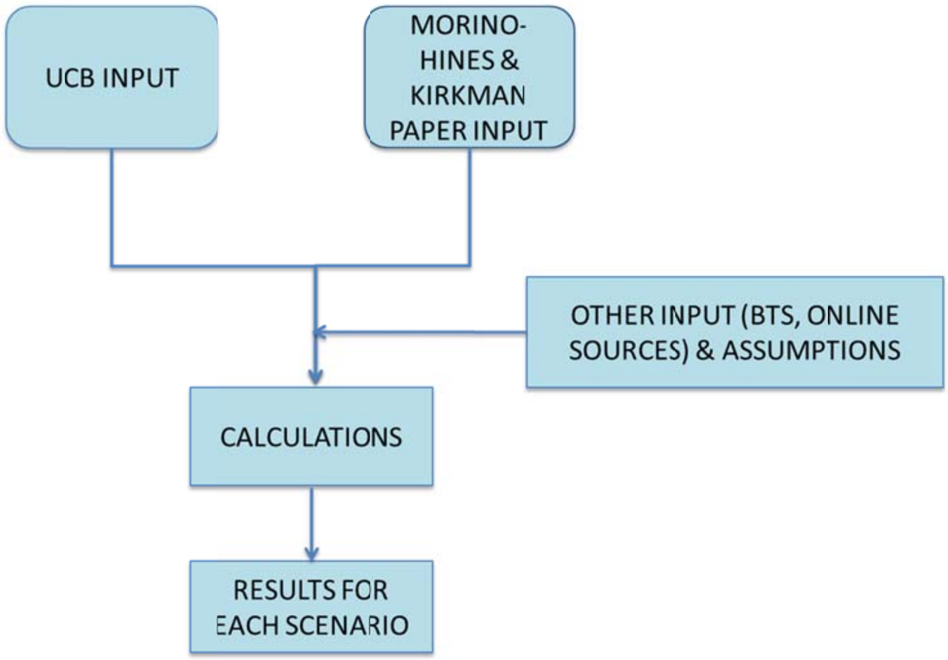

Figure 3-2: Benefits assessment methodology

The regression analysis conducted by Moreno-Hines and Kirkman (2013) included the estimation of coefficients that can be directly applied to the task of converting between average scheduled block time and certain dependent variables that, by themselves, are not monetized, but that can be monetized in a subsequent step using some benign assumptions. Table 3-1 below shows the coefficients used in this benefits assessment, as taken from that reference.

There are two limitations of that study that add some complication to the prospect of using the results in this benefits analysis. First, the empty cells in Table 3-1 represent coefficients that would be necessary for completeness, but were not included in the paper. In the paper, each model was developed in a parsimonious form, so independent variables that did not significantly improve the fit of the model were removed from the model specification. There is no way of knowing, at this point, what the un-estimated coefficients might have been, or what their p-values might have been.

Secondly, the set of carriers reported in that study does not match exactly the set used for the scenario generation as part of the scheduled block time impact modeling performed for this project. As a result, the team constructed a mapping between the two sets, which then necessitates the assumption that the effects to certain carriers are expected to be the same as those of other "similar" carriers. In particular, both studies used the major carriers American, Delta, and United, so those results were directly transferable. The Moreno-Hines and Kirkman study included Northwest Airlines, which was absorbed into Delta Airlines in 2010. In this study, we ignore the results for Northwest, although one could argue that their expected behavior might in some way manifest itself as part of Delta, but there is no way of quantifying this. Among low-cost carriers, the Moreno-Hines and Kirkman paper included American Eagle, JetBlue, and Southwest. In the current study, we consolidated low-cost carriers into one entity with respect to scenario generation. However, the major low-cost carriers differ noticeably in some statistics such as available seat miles that are used later in the benefits assessment process. Thus, we chose to distinguish between Airtran, JetBlue, and Southwest, using the same coefficients for all three (treated as the generic low-cost carrier in Table 2-5), but then applying their respective statistics for the other data.

Table 3-1: Block time benefits regression coefficients

Δ in Dependent Variable	**Δ in Explanatory Variable**	**AA**	**American Eagle**	**Delta**	**JetBlue**	**Northwest Airlines**	**Southwest Airlines**	**UA**	**US**
Average Scheduled Block Time	Average Actual Block Time	0.75	0.75	0.91	0.9	0.8	0.78	0.62	0.89
Average Daily Flights per Aircraft	Average Scheduled Block Time	-0.02		-0.02		-0.05		-0.02	
Average Pilot Salary per Available Seat-mile	Average Scheduled Block Time	6.4E-5				8.2E-5	6.7E-5		
Source: Moreno-Hines and Kirkman (2013)									

Table 3-2 below shows the data used for the benefits assessment. The Available Seat Miles (ASM), numbers of pilots/copilots, fleet size, and departures per year are from BTS. The wage data is also from BTS, specifically US DOT Form 41, Schedule P6 & P10. The yearly ownership cost per aircraft is computed as a weighted average. For each carrier, their fleet is stratified into different airframe types, each with different ownership costs. The average ownership cost per aircraft is then a weighted average of these values, weighted by the fraction of the total fleet represented by that particular airframe type. The data for this analysis are from Aviation Daily (2013).

The three following tables show the results of the analysis, one for each of the scenarios. In Scenario 1, American and Delta can save over 1.5 aircraft apiece, resulting in significant savings in ownership costs. The savings to United are more modest. As mentioned above, the necessary data to compute these savings for the low cost carriers are missing. For all of the carriers, however, it was possible to compute expected savings in pilot and flight attendant salaries that would be realized by conducting the same operational tempo with fewer total aircraft. Again, the numbers for American and Delta are high, and for United quite low. The savings for the LCCs are somewhere in between.

Table 3-2: Airline specific data for benefits analysis

				Low Cost Carriers (LCC)		
	American Airlines	**Delta Airlines**	**United Airlines**	**Airtran**	**JetBlue**	**Southwest**
Available Seat Miles (2011)	9.00E+10	1.09E+11	6.27E+10	2.34E+10	3.09E+10	1.03E+11
Pilots/Co-pilots (2011)	4898	6980	3731	1570	1730	5676
Departures per year (2011)	531,000	729,000	319,000	246,000	209,000	1,142,000
Mean Annual Wage for Pilots (2011)	$139,963	$150,099	$125,690	$128,225	$139,744	$203,196
Mean Annual Wage for Flight Attendants (2011)	$51,197	$40,475	$37,888	$32,088	$37,987	$54,120
Fleet Count	608	722	697	129	183	582
Yearly Ownership Cost per Aircraft	$1,766,492	$1,748,785	$2,366,392	$1,949,242	$1,735,440	$1,457,649

Table 3-3: Results for benefits Scenario 1

				Low Cost Carriers (LCC)		
	American Airlines	**Delta Airlines**	**United Airlines**	**Airtran**	**JetBlue**	**Southwest**
% of Saved Aircraft	1.61	1.56	0.30	0.00	0.00	0.00
Savings from Aircraft Saved	$17,326,246	$19,717,863	$4,998,482	$-	$-	$-
Average Reduction in Pilot salaries for current scenario (salary savings)	$11,116,924	$15,040,097	$762,955	$2,857,420	$3,769,531	$12,578,477
Average Reduction in Flight Attendant Salaries (salary savings)	$4,066,454	$4,055,643	$229,985	$715,063	$1,024,682	$3,350,200

Scenario 2 exhibited the most pronounced reduction in scheduled block times, and hence should produce the greatest expected savings. In Table 3-4 below, we can see that this is true. The same relative standings amongst the airlines holds as before, which is to be expected, because the cost coefficients are the same.

Table 3-4: Results for benefits Scenario 2

				Low Cost Carriers (LCC)		
	American Airlines	**Delta Airlines**	**United Airlines**	**Airtran**	**JetBlue**	**Southwest**
% of Saved Aircraft	3.77	2.81	0.43	0.00	0.00	0.00
Savings from Aircraft Saved	$40,487,757	$35,419,125	$7,103,105	$-	$-	$-
Average Reduction in Pilot salaries for current scenario (salary savings)	$25,977,889	$27,016,471	$1,084,199	$4,757,133	$6,275,648	$20,941,091
Average Reduction in Flight Attendant Salaries (salary savings)	$9,502,440	$7,285,136	$326,821	$1,190,461	$1,705,927	$5,577,530

Finally, Table 3-5 below shows the results for Scenario 3, which was representative of an intermediate level of reduction of scheduled block time.

Table 3-5: Results for benefits Scenario 3

				Low Cost Carriers (LCC)		
	American Airlines	**Delta Airlines**	**United Airlines**	**Airtran**	**JetBlue**	**Southwest**
% of Saved Aircraft	2.69	2.18	0.37	0.00	0.00	0.00
Savings from Aircraft Saved	$28,907,002	$27,568,494	$2,366,392	$-	$-	$-
Average Reduction in Pilot salaries for current scenario (salary savings)	$18,547,406	$21,028,284	$923,577	$3,736,626	$4,929,387	$16,448,777
Average Reduction in Flight Attendant Salaries (salary savings)	$6,784,447	$5,670,390	$278,403	$935,082	$1,339,969	$4,381,030

There are a few ways in which this analysis might be improved, with additional data resources. It is possible that some elements of ownership costs are not included (perhaps insurance, some maintenance costs, etc.). When determining the savings in aircraft ownership costs, it was assumed that any hypothetical drawdown in fleet size would follow the same fleet mix proportions as currently exist, while in reality a carrier would be very specific about which parts of its fleet it was ready to sell or otherwise retire. As mentioned several times above, the coefficients necessary to model aircraft savings for low cost carriers were not available, although perhaps those data could be acquired from the original Mitre team that did that study. Finally, this benefits analysis monetizes benefits from improvements in strategic flight predictability. There would also be benefits to be realized from improvements in tactical flight predictability, and these could be quantified as well.

4 Flight Predictability and Airline Fuel Cost

In this chapter, another aspect of the potential benefit of improved flight predictability is investigated. Different from the strategic flight predictability, in this chapter the flight predictability has operational impacts.

4.1 Background

It is important to understand the benefits of predictability and how these benefits can be monetized. Previous studies focus on the relationship between flight time dispersion and scheduled block time, which is a major airline cost driver. (Hao and Hansen, 2013; Wojcik and Mondoloni, 2013) In this study, we investigate an additional benefit mechanism from improved predictability—fuel savings. On first consideration, it may seem implausible that increasing predictability will affect fuel consumption. An automobile commute, after all, burns the same amount of gas in following a certain route with a certain speed profile, whether that speed profile is highly variable or very consistent from day to day. Why should a flight by an aircraft be any different?

The difference is fuel loading. While for automobiles this is a simple matter of occasionally filling up the tank, in aviation the choice of how much fuel to load for a flight, termed fuel uplift, is a delicate economic tradeoff. On one hand, airlines must load a sufficient quantity of fuel for flights to avoid any risk of fuel exhaustion and to reduce the likelihood of a fuel-related diversion. In practice, this means that considerably more fuel is loaded than is likely to be burned, and thus that most flights land with considerable fuel in their tanks. Loading this fuel, however, is not without penalty: aircraft fuel burn is highly sensitive to aircraft weight, and thus to the amount of fuel loaded since loaded fuel accounts for a large portion of the total aircraft weight.

Even the most stingy and courageous flier would be willing to pay for some extra fuel to ensure that a flight has enough fuel on board to complete its mission. Indeed, toward this end federal regulations stipulate minimum fuel reserves that must be boarded to each flight, and in some conditions also require sufficient fuel to fly to an alternate airport. In addition to reserve and alternate fuel, contingency fuel is also boarded. The amount of contingency fuel loaded is discretionary, and reflects the airline dispatcher's assessment of the "downside" risks that may lead to additional fuel burn beyond what is projected by the flight plan. Contingency fuel, together with the decision of carrying extra fuel to fly to an alternate airport when it is not required by federal regulation, thus represents the dispatcher's hedge against unpredictability. This, in combination with the effect of fuel uplift on fuel burn, suggests a connection between unpredictability and fuel consumption.

Recognizing the link between fuel uplift and predictability allows for both the monetization of predictability and the identification of new strategies for reducing aviation fuel consumption. There is intense focus on reducing fuel consumption from all stakeholders both to preserve the financial health of the airline industry and to minimize environmental impact. Airlines are moving aggressively to reduce fuel consumption because of rising fuel costs, which have gone from \$21 billion in 2009 to \$31 billion in 2012, and now account for 27% of airline operating costs, based on Bureau of Transportation Statistics data. (BTS, 2009; 2012) Higher fuel costs force airlines to increase their ticket prices, which in turn suppresses demand. With the tension in the Middle East and the increasing fuel demand from China, relief in fuel prices is not expected soon (Abdelghany and Abdelghany, 2005). Fluctuations in fuel price further exacerbate the problem. In 2008, jet fuel prices reached levels more than three times those of 2004, followed by a sharp decrease in 2009. In the future, climate change policies and environmental attitudes of potential air travelers may further increase or destabilize effective fuel prices (Ryerson et al., 2011). Thus, by economizing on fuel airlines reduce their exposure to an economic "wild card."

In this paper, we investigate the relationship between predictability and fuel uplift for a major US carrier. As discussed before, there is a direct physical relationship between fuel uplift and fuel consumption. Moreover, the fuel uplift determined by dispatchers may be affected by the flight time unpredictability. This might be the underlying mechanism to explain the results found by Ryerson et al. (2011) that when airlines add additional "buffer" time to flight schedules, fuel consumption increases. In this study, we exploit a large and recent flight level data set provided by a major US airline and merge this dataset with other publicly available datasets that incorporate NAS operating characteristics. The data included for each flight are the amount of loaded fuel; fuel burn and its rate; scheduled, planned, and actual flight times; and delays. Unpredictability of a flight can be captured in the variation (standard deviation) of flight time among all the flights between a specific OD pair, departure time bank and month. This data set enables us to evaluate different metrics for flight predictability and to estimate the relationship between predictability and fuel uplift, while controlling for other relevant factors such as terminal weather and traffic. Then by exploiting established relationships between fuel uplift and fuel consumption, we are able to evaluate the value of predictability in terms of fuel cost savings.

The remainder of this chapter is organized as follows: section 4.2 provides a review on recent literature regarding monetizing transportation disabilities and fuel saving practice; section 4.3 introduces the fuel loading practice from a representative US carrier; section 4.4 presents our modeling approach, wherein predictability is quantified and related to fuel uplift, and the estimation results; section 4.5 presents methods and results for estimating the cost to carry additional fuel loaded as a result of flight unpredictability; and section 4.6 provides a summary and conclusion.

4.2 Literature Review

Understanding the drivers of performance and how they are measured is a first step toward rationalizing system improvements; monetizing the performance metrics for use in a cost-benefit analysis is the next step for. Previous studies (on flight predictability?) focus on the relationship between flight time dispersion and scheduled block time, which is a major cost driver for airlines. The estimation of airline preference structures is particularly difficult because we only know what they reveal by their actions; while in truth the underlying decision-making process is highly complex (Mark you and Amy have a paper about this?) To estimate delay, additional fuel consumed due to additional operating time is summed with lost passenger time. The monetization of predictability is less straightforward. We must establish airline pre-flight planning practices that influence system cost in the presence and absence of a predictable airspace system. A critical pre-flight planning decision is the amount of fuel to place onboard. In fact, Ryerson et al. (2011) find that airlines that when airlines add additional "buffer" time to flight schedules, fuel consumption increases; the authors conclude that the additional fuel loaded in the planning phase adds weight to the aircraft, thus increasing fuel consumption. We propose to build on this concept and directly investigate an additional benefit mechanism from improved predictability—fuel savings due to decreased fuel loading.

Despite the important role fuel loading plays in fuel consumption and the delicate tradeoffs involved in determining fuel uplift, there is little literature specifically on fuel loading strategies in aviation. There are studies to reduce fuel consumption in various directions such as substituting all connecting flights in the US with non-stop flights (Jamin et al., 2004), substituting narrow body jets with turboprops (Ryerson and Hansen, 2010), and investigating Continuous Descent Approaches (CDA) and the Airspace Flow Programs (AFP) aiming at coordinating ground and air operations (Clarke et al., 2004). There is significant work investigating ground-based fuel savings measures such as single-engine taxi (Khadilkar et al., 2012) and delayed pushback procedures (Simaiakis, 2012). In an effort to investigate how different components of delay impact fuel consumption, Ryerson et al. (2011) use simulated and actual airline fuel consumption data and find that fuel attributed to planned delays (in the form of schedule padding) accounts for about 20% of the fuel that can be attributed to unplanned delays. They estimate that one minute of airborne delay burns 50-60 lbs of fuel, compared with 4.5-12 lbs of fuel for a minute of schedule padding and 2.3-4.6 lbs for a minute of departure delay, for two aircraft types. Building on the findings of Ryerson et al. (2011) that adding a buffer to scheduled operations, which is a reflection of historical unpredictability, leads to increased fuel consumption, in this paper we investigate the underlying relationship between predictability and fuel uplift for a major US carrier.

4.3 Fuel Loading Practice

The determination of fuel uplift for a specific flight is an important and safety-critical aspect of airline flight planning. During the flight planning phase, flight plans are created by dispatchers at an airline control center. Each flight plan is documented in a flight release, which is created around two hours prior to departure time. Each dispatcher typically works a 9 or 10 hour shift during which, for domestic dispatchers, about 40 flights will be planned. The flights for a particular dispatcher's desk are typically organized by geographic region. Dispatchers typically work the same desk, and thus the same set of flights, from day to day. In addition to flight planning, dispatchers perform other duties to ensure the safe operation of a flight from origin to destination. These include providing pilots with real-time updates, coordinating between various parties to resolve maintenance issues, and continuously monitoring the flight from takeoff to landing.

The information typically considered by dispatchers when planning flights include current and forecast weather conditions at the origin, destination, and en route, restrictions or notifications from air traffic control, and specific flight routings. The major decisions that affect fuel loading include the amount of taxi fuel, the amount of contingency fuel, and the choice of alternate destination airports. The taxi fuel and contingency fuel are chosen in a continuous fashion to account for additional fuel that might be needed at some point along the mission. The fuel burn rate of each flight is different due to differences in flight length, aircraft type, winds, payload, and a number of other factors. Thus, dispatchers use a standardized metric for the amount of fuel loaded onto a flight, which is the number of minutes of flight time at specific operating conditions. This allows dispatchers to think in terms of flight time uncertainty rather than fuel burn uncertainty,

A dispatcher must determine the amount of contingency fuel to be uplifted. This is an amount of fuel that is to be used in the case of unexpected conditions during the flight, such as additional time on the ground or in the air from unexpected delays, re-routing, flight level changes or airborne holding. It is typically measured in minutes based on the fuel burn rate of the aircraft in normal cruise conditions. Airline policy may dictate a minimum amount of contingency fuel for a domestic flight (for example, 10-15 minutes) regardless of flight conditions, but dispatchers usually add more. Contingency fuel is, in essence, a reflection of expected operational degradation, or unpredictability. Dispatchers may be presented with guidance regarding the historical distribution of actual fuel burn relative to planned fuel burn for similar flights to help them determine contingency fuel. In practice, however, dispatchers uplift more contingency fuel than even the far right tail of this distribution would suggest, in order to provide extra protection against unforeseen circumstances. Contingency fuel is one of four fuel components: the others are mission fuel, reserve fuel, and alternate fuel. Mission fuel is the amount of fuel needed to complete a flight, based on distance, planned route, and aircraft type-specific fuel burn rate. Federal rules exist regarding the amount of reserve fuel that must be carried. U.S. Federal Aviation Regulations (14 C.F.R. § 91) require that a commercial flight must have enough fuel to complete the flight to the intended destination airport, fly from the destination airport to the alternate airport (if required), and then hold in the air for 45 minutes at normal cruising speed (Ryerson and Churchill, 2013).

In addition to choosing the amount of contingency fuel and taxi fuel for a particular flight, dispatchers evaluate whether an alternate destination airport is needed. Certain weather conditions require the addition of an alternate. According to the FAA regulation, the weather conditions requiring an alternate are less than 3 miles visibility and less than 2000 ft ceiling for the destination airport at the Estimated Time of Arrival (ETA) ± 1 hour. If the destination is equipped for CAT 1, 2, or 3 Instrument Landing System (ILS) operations, then the required minimums for an alternate are reduced slightly. Although thunderstorms are highly disruptive to airport operations during summer times, they are not part of the FAA's alternate policy. However, airlines may have their own internal policy, such as planning for an alternate when thunderstorms are forecast for ETA ± 1 hour. Many alternate airports are listed even when these conditions are not met. Thunderstorm forecasts can vary in terms of probability, and in practice (as observed by the authors) we see that any chance of a thunderstorm typically results in an alternate airport listed on the flight release. Also, an alternate airport is sometimes listed even when the weather conditions do not require it and no thunderstorms are forecast; this is similar to adding contingency fuel. Likewise, sometimes a second alternate airport is added to the flight plan, although this is never required either by FAA or by airline fuel policy. This usually happens when the first alternate airport has marginal weather and the dispatcher doesn't feel fully secure by only having one alternate. The alternate fuel is determined by the farther of the two alternates, so a second alternate does not necessarily increase the fuel load for a particular flight. The additional fuel loaded due to alternate airports is based on the distance between the original destination and the alternate airport, as well as the aircraft type.

4.4 Data and Modeling

4.4.1 Data Collection

To estimate the impact of unpredictability on fuel uplift, data were collected from three sources: fuel and flight statistics from a major United States-based air carrier, weather information from the National Oceanic and Atmospheric Administration (NOAA), and airport data from the FAA Aviation system Performance Metrics (ASPM) database.

The US carrier used in our analysis operates an extensive domestic and international network serving all continents except Antarctica. Together with its subsidiaries, the airline operates over 5,000 flights every day and has approximately 80,000 employees. The dataset provided from the airline includes all domestic flights between April 2012 and May 2013, inclusive. There are altogether 562,274 flights during the 14 months for which data were collected. The airline dataset contains flight-by-flight data on planned and actual fuel consumption, fuel uplift in all categories (taxi, contingency, and alternate), in units of minutes, as well as flight information such as equipment, origin and destination, planned and actual flight times (including out-off-on-in times), and delay information. It also provides actual fuel burn data by flight phase, including taxi-out, wheels-off to wheels-on, and taxi-in.

The weather data collected from the NOAA database includes both the actual weather and national weather forecast (TAFs) information for major US airports. The actual and forecast weather information contains ceiling, visibility as well as indicators of the presence of thunderstorms, snows, VFR/IFR conditions in an hourly manner for each day and each airport. VFR / IFR refers to either visual flight rules or instrument flight rules and is an indicator of overall favorable or unfavorable terminal flight conditions, respectively. The weather data were matched with the flight-level airline dataset to recreate the conditions seen by dispatchers during the time of flight planning. We used two hours prior to the scheduled time of departure as a proxy for the time of flight plan decisions made by the dispatchers. At this time, we found the actual weather conditions at the origin and destination airports, as the actual weather at time of flight planning is also found to be influential on dispatchers' fueling decisions. Dispatchers should not be affected by the on-going weather, however, most dispatchers claim that they also refer to the actual weather during flight planning, as an extra reassurance. For the weather forecasts, we used the most recent forecast that was issued for the origin and destination airport at two hours prior to scheduled departure time, and found the weather conditions that were forecast for the scheduled time of departure at the origin and the scheduled time of arrival at the destination. These weather variables are consistent with the flight planning process revealed by our on-site observations and discussions with flight dispatchers.

The FAA ASPM database includes quarter hourly data for the 77 large airports in the US on arrival traffic conditions. It contains average arrival delay by the quarter hour for each airport to depict level of congestion at the airport.

The three datasets are merged in a manner that will be described below. After merging and some filtering to keep the dataset robust that will be explained shortly, there were 221,893 flights in the dataset during the 13-month time period. Aggregately, there were 3,227 groups of combination of OD pair, month and dispatcher work shift, denoted as *od_month_shift.*

4.4.2 Estimation Methodology

We seek to statistically estimate the contribution of flight unpredictability to contingency fuel uplift. As noted above, in addition to loading contingency fuel per se, dispatchers sometimes add fuel by adding alternates. We therefore estimate two separate models: one with contingency fuel uplift as the dependent variable, and the other with contingency plus alternate fuel uplift as the dependent variable. As explained above, in both cases we express this fuel quantity in minutes, which is a common practice in fuel loading. We denote by $CF(od, m, s, dest, d, h, q, dis)$ the contingency fuel uplifted on the individual flight between OD pair *od* on date *d* at hour h and quarter hour q. Hour h could be the departure and arrival hour for the flight, respectively for the purpose of our data merging that will be explained later. The empirical definition of predictability (discussed below) will also require including the month *m* and shift *s* when the flight departs as arguments. We divide the departure time of the flight into three shifts in a day: 5am to 3pm as the morning shift, 3pm to 10pm as the afternoon shift, and 10pm to 5am as the midnight shift. These shifts are the actual shifts on dispatchers' work schedules. The distribution of shifts relies mainly on a seniority basis. To capture variation across dispatchers' fuel uplift practices we also index on the identity of the dispatcher who planned this specific flight, denoted as *dis*. Similarly, we define the variable $TOT(od, m, s, dest, d, h, q, dis)$, which is the sum of contingency and alternate fuel, reflecting the total amount of fuel uplift for contingencies/unplanned events.

The variables $CF(\cdot)$ and $TOT(\cdot)$ are the dependent variables in our models. The independent variables will capture flight predictability and other variables that affect contingency and alternate fuel uplift, such as weather and traffic demand. As a metric for predictability, we use the standard deviation of actual airborne time for flights serving the same *od* pair with a departure time in the same shift s taking place in the same month m. We segment flights in this manner based on interviews with the airline dispatchers who indicated that fuel loading judgment is most greatly impacted by month, shift, and *od* pair. In contrast, the aircraft type and specific flight number are not specifically considered because most dispatchers do not pay attention to these details in fuel loading decisions. The calculation of the standard deviation of actual airborne time takes place on these aggregated sets of flights segmented by *m*, *s*, and *od*. This is denoted as $stdair(od, m, s)$. To ensure robustness, only combinations of these arguments with more than 25 flights are kept in our dataset. By examining the effect of dispersion in historical flight performance, we seek to capture the impact of the unpredictability in flight time on dispatcher uplift of contingency fuel and alternate fuel. We also include the associated mean value $avgdair(od, m, s)$ to capture uncertainty that arises from increasing average flight duration, which means results in less reliable forecasts of conditions en route and at the destination airport. Note that only airborne time is considered, because it is the largest component of the total flight time and an even more dominant source of fuel burn.

In addition to the dispersion of airborne time, the difference between actual airborne time and planned airborne is also a reflection of unpredictability. In everyday operation, the actual airborne time is sometimes different from the planned airborne time due to unforeseen en route conditions. Thus in this model, we include the mean and standard deviation of the difference in airborne time from flight plan. The variable $dif = actual\ airborne\ time - planned\ airborne\ time$ is calculated for each flight, and used to calculate the mean and standard deviation $avgdif(od, m, s)$ and $stdif(od, m, s)$, with the same level of aggregation.

We assume that dispatchers consider experience in the recent past when making fuel uplift decisions. Thus we specify our models so that decisions for flights taking place in month *m* are based on the above flight performance metrics for the previous month *m-1*.

As noted in section 2, the addition of an alternate on a flight plan, and the related fuel consumption, may be mandated by federal regulations due to certain weather conditions. In an effort to separate the impact of predictability and the presence of weather, we define variables to capture the weather conditions that trigger the mandated addition of an alternate. The weather information includes two parts: the actual and forecast information for both the origin and destination airport. Even though dispatchers are only supposed to base their fuel loading decisions on the forecast weather information during the departure and arrival time of the flight, we discovered that dispatchers tend to rely on current weather condition at the time of their flight planning for fuel loading decisions as well through our observation and interview with dispatchers in the airline's control center. This over-cautious behavior is also included in the model through actual weather information. Moreover, the weather conditions at both the destination airport (denoted as *dest*) and the origin airport (denoted as *orig*) are included. They affect both the taxi and air-borne time of the flight and thus the fuel loading consideration.

Dispatchers typically make the fuel loading decisions around the time the flight plan is created. Small deviations in this time do occur from flight to flight, but typically these decisions are made 2 hours prior to the flight's scheduled departure time; thus this can also be called the dispatch time. Therefore, we merge the weather data with the flight-level data to recreate the real-time and forecast weather that was available at the dispatch time for each flight. For the real-time weather, we find the actual weather at the origin and destination at the dispatch time. For the forecast weather, first we have to find the most recent forecast that was issued prior to the dispatch time and refer to the forecast conditions for the origin at the planned departure time and for the destination at the planned arrival time.

For the forecast weather at destination airport *dest* (origin airport *orig*), the hourly NOAA data is merged to the realized flight by the hour *h* in which the flight is planned to arrive (depart) in the flight plan. Variable $lowc\_F(dest, d, h)$ is 1 if the TAFs forecast ceiling at airport *dest* in day *d* and hour h in which the flight is planned to arrive, is lower than 2000 feet, and 0 otherwise. Similarly, $lowv\_F(dest, d, h)$ is 1 if the forecast visibility at airport *dest* in day d and hour h is lower than 3 nautical miles, and 0 otherwise. These criteria for low ceiling and low visibility are the threshold in the federal regulation for alternate requirement. To further capture overall conditions, the variable $IFR(dest, d, h)$ is 1 if the airport *dest* has IFR conditions forecast in hour *h* of day d, and 0 otherwise. The NOAA weather also provides dummy variables indicating snow and thunderstorm at the airport. Variable $snow\_F(dest, d, h)$ ($Tstorm\_F(dest, d, h)$) is 1 if snow (thunderstorm) is forecast at airport *dest* in day d and hour h, and 0 otherwise. For forecast weather, the variables depicting conditions at origin airports are the same as the destination variables, with origin airports denoted as $orig$. In addition to the forecast weather condition, the actual weather conditions at dispatch time are also included. The variables are similar to the forecast ones, without the suffix $\_F$. The actual weather variables are merged to the flights by the time of dispatching, which is two hour prior to the planned departure time of the flight.

Lastly, to capture the effect of congestion at the destination airport on contingency and alternate fuel uplift, we include a variable depicting arrival delay at the destination airport. Variable $Arr(dest, d, q)$ is the average arrival delay per flight in minutes at destination airport *dest* in day d and quarter hour q. The quarter hourly average arrival delay information is from the ASPM dataset.

As the airline dataset includes flight-specific information regarding the dispatcher identity and there exists substantial variation in fuel loading behavior across dispatchers, fixed effects for individual dispatchers are included in the model. Each flight has a specific dispatcher *dis*, with the fixed effect captured in variable ρ_{dis}. We additionally define a fixed effect for each month. The purpose of this variable is to capture seasonality effects not captured explicitly by our terminal congestion and weather variables. For example, en route weather conditions are highly subject to seasonality and greatly influence fuel loading decisions. However, the en route weather conditions are not captured in the terminal weather variables that we include in the model. The monthly fixed effects are denoted as γ_m.

The model formulations are as follows. Equation 1 models the behavior of loading contingency fuel only, while equation 2 explains the behavior of loading the total contingency and alternate fuel:

$$
\begin{aligned}
&\{CF(od, m, s, dest, d, h, q, dis), TOT(od, m, s, dest, d, h, q, dis)\} \\
&= \alpha_0 + \alpha_1 \times avgair(od, m-1, s) + \alpha_2 \times stdair(od, m-1, s) + \alpha_3 \times avgdif(od, m-1, s) + \alpha_4 \times stdif(od, m-1, s) \\
&+ \sum_{i \in \{dest, orig\}} (\beta_{1i} \times lowc(i, d, h) + \beta_{2i} \times lowv(i, d, h) + \beta_{3i} \times IFR(i, d, h) + \beta_{4i} \times snow(i, d, h) + \beta_{5i} \times tstorm(i, d, h)) \\
&+ \sum_{i \in \{dest, orig\}} (\lambda_{1i} \times lowc\_F(i, d, h) + \lambda_{2i} \times lowv\_F(i, d, h) + \lambda_{3i} \times IFR\_F(i, d, h) + \lambda_{4i} \times snow\_F(i, d, h) + \lambda_{5i} \times tstorm\_F(i, d, h)) \\
&+ \pi \times Arr(dest, d, q) + \rho_{dis} + \gamma_m + \varepsilon
\end{aligned}
$$

4.4.3 Estimation Results

The estimation results are shown in Table 4-1. Note that because our assumption is that the fuel uplift is impacted by the flight performance data of the previous month, the model is estimated on 13 of the 14 months from the data. The left portion of Table 1 shows the estimation results for Eq. (1), for which the dependent variable is contingency fuel in minutes. The top part of the table shows the intercept and the predictability metrics. The intercept is around 20 minutes of contingency fuel. This number matched our expectations based on our observations and a company policy of 10-15 minutes minimum of contingency fuel. Since the weather variables are capturing adverse conditions, the intercept can be interpreted as what the fuel uplift would be under good weather conditions. Average airborne time over the past month has a small positive coefficient, indicating that longer flights are loaded with more contingency fuel than shorter flights. However the marginal increase in loaded contingency fuel for 1-minute increase in average airborne time is quite small. The coefficient on the standard deviation of airborne time is 0.88, indicating that one minute of variation in the airborne time in previous month will lead to the uplift of an average of almost one minute of additional contingency fuel. This indicates that an increase in the unpredictability of the daily flight operation, based on our metric, will lead to an increase in contingency fuel uplift at an almost one-minute-to-one-minute ratio. The deviations captured in variables *avgdif* and *stdif* both have significant coefficients as well. Variable *avgdif* has a small negative coefficient, indicating that an increase in the average disparity between planned and actual flight time decreases fuel uplift. This might seem unreasonable at first. However, since the weather variables are also capturing some of the causes of flight time deviation, this effect might be absorbed in the weather variables, resulting in a negative coefficient for the average deviation itself. The second-moment metric (standard deviation) has a positive impact on contingency fuel loading, indicating it increases fuel uplift. The first-moment metric has slightly smaller coefficient than the second-moment metric, but both are on a very small scale. Also, it is worth noting that the standard deviation of airborne time has a much larger coefficient than the standard deviation of the difference between actual and planned trip time. This suggests that dispatchers are more strongly influenced by overall variability in flight time than by the accuracy of the predictions of the flight planning system, although in principal, it should be the latter that receives greater consideration.

The middle part of the table shows the estimates for the weather variables, separated by destination and origin airport. Most weather dummy variables have a positive coefficient, indicating that adverse weather at the airports will increase contingency fuel uplift. Thunderstorms have the largest impact and low ceiling the smallest. Overall, forecast weather has a larger impact on the uplift of contingency fuel than the actual weather, indicating that dispatchers' fueling decisions are more affected by the weather forecast than the actual weather at the time of flight planning. However the actual weather conditions still have significant effects, validating our observation on dispatcher behavior. Moreover, the weather conditions at the destination airport, both actual and forecast, are greater contributors to contingency fuel than the weather conditions at the origin airport. As the arrival is the more unpredictable phase of the flight and could induce a large amount of fuel consumption from holding in the arrival queue due to adverse weather, these results are consistent with our expectations. Lastly, regarding the traffic conditions at the destination airport, a 1-minute increase in the average arrival delay at the destination will lead to a 0.14 minute increase in the contingency fuel loaded. Arrival delay can derive from arrival queuing at the destination airport, which may lead to airborne holding, but it may also derive from gate delay at the origin, which has little impact on fuel burn. This probably explains why the arrival delay coefficient is considerably less than 1.

Table 4-1: Estimation Results for Fuel Uplift Models

	Contingency min		**Total min**	
Variable	***Coefficient***	***P-value***	***Estimate***	***P-value***
Intercept	20.601	<.0001	20.883	<.0001
avgair	0.0153	<.0001	0.0103	<.0001
stdair	0.883	<.0001	1.657	<.0001
avgdif	-0.0279	<.0001	0.0938	<.0001
stdif	0.0096	<.0001	0.0361	<.0001
lowc_dest	-0.0377	0.7191	10.399	<.0001

lowc_dest_forecast	1.686	<.0001	14.057	<.0001
lowv_dest	1.858	<.0001	16.041	<.0001
lowv_dest_forecast	3.483	<.0001	24.774	<.0001
tstorm_dest	3.267	<.0001	22.457	<.0001
tstorm_dest_forecast	8.875	<.0001	51.952	<.0001
snow_dest	1.0046	<.0001	11.896	<.0001
snow_dest_forecast	3.322	<.0001	26.517	<.0001
IFR_dest	0.658	<.0001	16.886	<.0001
Arrival delay_dest	0.136	<.0001	0.514	<.0001
lowc_ori	-0.145	0.1785	-0.0621	0.8434
lowc_ori_forecast	0.0527	0.6710	-0.672	0.0635
lowv_ori	0.470	0.0002	0.175	0.6297
lowv_ori_forecast	0.793	<.0001	2.023	<.0001
tstorm_orig	1.625	<.0001	1.824	0.0188
tstorm_orig_forecast	5.627	<.0001	8.381	<.0001
snow_orig	0.734	0.0002	-1.135	0.0497
snow_orig_forecast	1.263	<.0001	0.0933	0.8536
IFR_orig	0.643	<.0001	1.997	<.0001
month 1	-1.279	<.0001	-4.679	<.0001
month 2	-0.631	<.0001	-3.080	<.0001
month 3	-1.219	<.0001	-8.636	<.0001
month 4	1.018	<.0001	-1.784	<.0001
month 5	1.068	<.0001	1.633	<.0001
month 6	2.037	<.0001	2.811	<.0001
month 7	5.079	<.0001	14.467	<.0001

month 8	3.559	<.0001	9.974	<.0001
month 9	1.9999	<.0001	3.799	<.0001
month 10	0.912	<.0001	0.131	0.6888
month 11	-1.785	<.0001	-8.920	<.0001
month 12	0.000	-	0.000	
R-squared	0.2623		0.4163	

The right-hand columns of Table 4-1 include the estimation results using the sum of contingency minutes and the alternate fuel in minutes as the dependent variable. The estimation results are generally similar to the contingency fuel model on the left-hand side. The intercept of 20 minutes is similar to the intercept from the contingency fuel model. Since most adverse weather conditions are considered, the intercept depicts a typical amount of fuel uplift for a good weather day. Therefore, it is reasonable to see a similar intercept for the two models. Most other coefficients are larger than the left-hand model, as the dependent variable is larger in most cases. The coefficient estimate for standard deviation of airborne time is twice as large as the coefficient from the contingency-only model, implying that dispatchers uplift approximately 1.66 minutes more fuel for a 1-minute increase in the standard deviation of airborne time. This suggests that alternate fuel is a major component of dispatchers' hedge against uncertainty. Again the impact of airborne time unpredictability is much larger than the impact of flight plan unreliability, the mean and standard deviation of which now both have positive, albeit very small, impacts.

Since the dependent variable now includes alternate fuel, the weather conditions at the destination airport have a more dominant impact. The coefficients for all the destination airport weather variables are significant, for both actual and forecast weather, and are around 9 times larger in magnitude than the corresponding coefficients from the contingency-only model. For origin airports, the coefficients are similar to those in the contingency-only model, and some variables are not statistically significant, such as real-time low visibility and low ceiling.

The bottom part of Table 4-1 shows the monthly fixed effects. For ease of presentation the estimates of the dispatcher fixed effects are excluded. December is chosen as the baseline month to which all other months are compared. We see that summer months, especially July and August, have larger coefficients than other months. These two months are during thunderstorm season, which greatly impacts contingency and alternate fuel uplift. Although the adverse weather variables, in particular the thunderstorm variables, are included to account for the impact of convective weather, there are still effects from en route thunderstorms and other adverse weather conditions sensitive to seasonality that cannot be captured by the airport weather variables. The absolute difference in contingency fuel due to monthly fixed effects is almost 7 minutes, with the largest fuel load in July and the smallest in November. If we consider alternate fuel as well, the scale is much larger, ranging from -9 to 15 minutes for November and July, respectively, but the trend is much the same.

4.5 Cost to Carry Analysis

The previous section established that unpredictability in airborne time will lead to an increase in fuel uplift. The significance of this additional uplift is both financial and environmental. The rate of fuel consumption increases with weight; said another way, you spend fuel to carry fuel. There is a measurable *cost to carry*, or the pounds of fuel consumed per pound of fuel carried per mile. This rate varies across aircraft types and flight lengths with a general rule of thumb being that it costs about one-quarter to one half-pound of fuel to carry a pound of fuel (Leigh, 1995). There is therefore an additional amount of fuel consumed that can be attributed to the additional contingency and alternate fuel uplifted as a result of unpredictability. In this section, we quantify this added fuel and then translate it into fuel consumption, and then into costs in terms of purchase expense and emissions of carbon dioxide (CO_2), the most abundant of the Greenhouse Gases (GHG) contributing to climate change.

In an aviation system with no operational unpredictability, the standard deviation in airborne time would be zero. It follows that the coefficient α_2 on the variable $stdair(od, m, s)$ represents the fuel penalty of unpredictability. For the 221,893 realized flights during the 13 months for which data are collected, there are 3,227 groups of *od_month_shift*. For each flight observation, we calculate $\alpha_2 * stdair(od, m-1, s)$ as the contribution of variation in airborne time to the loaded fuel of this flight. The average contribution of the total 221,893 flights is then calculated. In a perfect scenario where no variation in airborne time exists, the loaded contingency fuel would be reduced by 6.12 minutes per flight. If we consider the sum of contingency fuel and alternate fuel, the reduction would be 11.28 minutes per flight. Again the loaded fuel is in the unit of minutes, which is a common practice in flight planning. While a perfect predictability scenario might include other differences, such as the elimination of differences between planned and actual flight times, we ignore those here because their impacts are small compared to that of standard deviation.

To estimate the savings in cost to carry from perfect predictability, we translate the excess minutes of contingency and alternate fuel of each individual flight into pounds of fuel using the fuel consumption per minute rates provided by the airline. These rates are specific to a particular flight, based on information in the flight plan such as equipment type and weather conditions. We next translate this into a quantity of fuel burned due to the loading of additional fuel using the airline cost to carry rates in units of lb per lb per mile. The results are presented in Table 4-2 in four categories. The first is the average fuel consumed (in lbs) due to additional uplift per operation. The second is the total amount of additional fuel consumed (in lbs) over the entire set of flights for which data are available. Due to reporting difficulties and the manual method some aircraft require for fuel reporting, the master airline dataset covers about 80% of the total operations. As such, the third category is the total amount of additional fuel consumed, in lbs, across the airline. We collected monthly domestic flight counts from the BTS T-100 Segment Database and extrapolated our results to these monthly counts. The fourth category is the total amount of additional fuel consumed (in lbs) across all airlines for all domestic operations. We collected domestic operational counts for all US carriers (those with \$20M or more in revenue per year) from the BTS T-100 Segment Database and extrapolated our results to this operational count again on a monthly basis.

The fuel quantified in the first two rows of Table 4-2 can be translated into cost and environmental externalities of fuel consumption. Reducing fuel consumption is a major initiative of the aviation industry as a whole. It is a way to reduce costs, and environmental impacts particularly related to climate change, manage the risk related to fuel price fluctuations and uncertainty surrounding a future environmental policy, and improve consumer perceptions of "greenness". As such there are many initiatives being considered in the form of policies, operational changes, and technology deployments. These ranges come from airline driven changes such as emphasizing single engine taxi procedures (Simaiakis and Balakrishnan, 2010; Nikoleris et al., 2011) and the federally-driven Next Generation Air Transportation System (NextGen) which promises significant fuel consumption reduction. We can translate the fuel savings from increased predictability into costs in terms of fuel prices and CO_2 emissions. As fuel prices fluctuate throughout the year and airlines have their own fuel contracts that may change the fuel cost they see, we estimate the cost to carry for three fuel prices: the average for the study year, about $3.00/gallon; and plus/minus $1.00/gallon (Airlines for America, 2013). To convert excess fuel into lbs of CO_2, we utilize the U.S. Environmental Protection Agency conversion factor for Jet Fuel (EPA, 2013). The results are presented in Table 4-2.

Table 4-2: Annual Cost to Carry Additional Contingency and Alternate Fuel from for Unpredictability in Terms of Fuel lbs, Fuel cost, and CO_2

Metric		Mean per Operation	Sum over Operations in the Dataset	Sum Extrapolated Over the Airline	Extrapolated over all Domestic Operations
Fuel (lbs)	**Contingency**	48.35	$1.07*10^7$	$3.56E*10^7$	$2.25*10^8$
	Contingency and Alternate	90.73	$2.01*10^7$	$6.69*10^7$	$4.23*10^8$
CO_2 (lbs)	**Contingency**	155.11	$3.44*10^7$	$1.14*10^8$	$7.21*10^8$
	Contingency and Alternate	291.07	$6.46*10^8$	$2.14*10^8$	$1.35*10^9$
Fuel cost at $2/gallon	**Contingency**	$14.43	$3.20*	$10.64*	$67.28*
	Contingency and Alternate	$27.08	$14.41*	$19.96*	$126.21*
Fuel cost at $3/gallon	**Contingency**	$21.65	$11.07*	$15.95*	$100.85*
	Contingency and Alternate	$40.62	$21.62*	$29.94*	$189.31*
Fuel cost at $4/gallon	**Contingency**	$28.86	$14.76*	$21.27*	$134.49*
	Contingency and Alternate	$54.17	$28.82*	$39.91*	$252.35*

*: in million dollars

The results in Table 4-2 provide us with the value, in terms of monetary costs and environmental externalities, of predictability. On a per flight basis, this value is $14.43 - $54.17 depending on fuel prices and whether alternate fuel is considered. Across all domestic flights, this value ranges from $67.3 – $252.4 million per year. To put these results in perspective, we first consider that in 2011 (a close proxy for our time frame) the total amount of jet fuel consumed was 12.1 billion gallons (EPA, 2013b). Therefore, the total amount of fuel consumed due to the lack of predictability in the system is about 1%. One percent may seem like a small number, however, it is in line with current initiatives branded as fuel saving “green” initiatives. Consider that the 1% translates to about 50-100 lbs per flight burned due to excess uplift (Table 4-2). While this number seems small, consider that the average flight in our sample consumes 553 lbs of fuel in taxi out and 233 lbs of fuel in taxi in, and the numerous efforts and research are taking place to reduce fuel consumption on the ground. It is also comparable to the savings estimated from use a continuous decent approach as compared to a conventional step-down approach (Cao et al., 2011). Efforts to reduce taxi fuel consumption and fuel consumption in descent involve investment, institutional change management, and the cooperation of federal, state, local, and private stakeholders as well as the traveling public. In showing that improved predictability has a value of fuel savings commensurate with other initiatives, we rationalize the consideration of investments that will improve predictability as a fuel saving and “green” measure.

4.6 Summary

In this paper, the relationship between flight predictability and contingency fuel loading is studied using detailed empirical data. Flight predictability is mainly depicted by the variation of airborne time because that most strongly affects dispatchers’ decision on contingency fuel loading. Flights are grouped by OD pair, departure time of day and month based on our observation of dispatchers’ fuel loading behavior. We found 1 minute of standard deviation of airborne time would lead to an additional 0.88 minutes of contingency fuel loaded to each flight. If we also consider alternate fuel, there would be a 1.66 minute increase in the sum of contingency and alternate fuel for a 1 minute increase in standard deviation of airborne time. Our other findings show that the deviation from planned flight time has a much smaller but significant impact on loaded fuel than overall airborne time variability, indicating dispatchers are more influenced by overall variability rather than flight plan accuracy; the forecast weather at the destination airport is the most influential weather factor for fuel loading, among which thunderstorm is the largest contributor, and low ceiling is the smallest. Also, there are significant seasonal fixed effects, with dispatchers loading fuel more fuel in the summer (July and August), probably to account for en-route thunderstorms and other weather factors not included in in our model.

To further quantify the impact of flight predictability on fuel loading, we calculate the value of flight predictability assuming a hypothetical scenario where there is no variability of airborne time for flights in a given OD pair, shift, and month. If the standard deviation of airborne time for all the flights are zero, on average there is a reduction of 6.12 minute per flight of contingency fuel, and 11.28 minute per flight of the sum of contingency and alternate fuel. This extra boarded fuel requires additional fuel to carry. Based on our calculation, for an average flight 48.35 lbs of fuel is consumed to carry the extra contingency fuel and 90.73 lbs to carry the extra contingency and alternate fuel that results from flight time variability. This translates into a cost to US domestic airlines on the order of $120.55 – $452.43 million per year. Social costs from additional emissions of GHG, not explicitly estimated here, add to this total value. Of course, it is not realistic to assume that all variability in flight times can be eliminated. The figures presented should be viewed as a heretofore unrecognized benefits pool from increasing predictability. Individual projects large or small that improve predictability can draw from this pool, and in some cases this may tip the balance for the project business case. For purposes of comparison, a recent study estimates that flight delay, which likewise can never the eliminated and but is the major motivation for NEXTGEN, costs the US airline industry around $10 billion per year (Zou and Hansen, 2012).

Our analysis establishes a behavioral link between flight time variability and fuel loading. Given that link, we have estimated the potential cost and emission savings from reducing variability.

With our analysis, the FAA's proposed study on flight predictability is provided with a more complete benefit motivation. The improvement in flight operational predictability will benefit not only the operational performance, but also the airlines' long-term fuel cost, which is rather the strategic planning aspect. Neglecting the less obvious benefit manifested in this study would lead to underestimating benefit in developing the predictability study. Another way to attain these savings is to change dispatcher behavior. It is not obvious that there is a sound operational reason to load more contingency fuel because flight time varies, since much of this variation is captured in the flight plan. Whether, and how, it may be possible to persuade dispatchers to attenuate the influence of flight time variability in their fuel loading decisions is a subject of ongoing research.

5 Predictability of Ground Delay Program Called Rates

An aspirational goal of the project had been to deduce something quantitative about the predictability of parameters of traffic management initatives (TMIs) invoked by the FAA to respond to unforeseen imbalances between arrival and capacity on a day of operations. In particular we wanted to look at ground delay programs (GDPs), and their called capacity rates, which end up affecting the amount of ground delay assigned to each flight. This assignment of delays also affects carriers' decisions to cancel flights.

The problem with quantifying parameters related to GDPs is that they are extremely complicated, with numerous contributing factors, and finding statistically justifiable patterns in the data can be a quite elusive goal. Because it was not a primary goal of this project, this effort was never more than gestational; however, some preliminary data gleaned from real GDPs might serve to illustrate both the complexity of the problem and perhaps give some insights as to starting points for later projects where this task can be a more central focus.

5.1 Comparison of ASPM vs ASDE-X Data for 4 Days in BOS

A first approach to look in more depth at the predictability of call rates was to compare the actual departures and arrivals as reported in ASPM and ASDE-X for 4 days in BOS. The point with this analysis is to demonstrate that it is not even straightforward to determine from publicly available data sets what actually happened on a given day, let alone make inferences about what predictions and expectations might have been. We used ASDE-X data for 4 days in 2010 (June 10th, July 3rd, July 13th and October 5th) and data retrieved from ASPM for the same days. In the following figures we can see how these two different sources of data compare. In each figure we have included also the claimed (by ASPM) airport departure rate (ADR) and airport arrival rate (AAR). In Figure 5-1 and Figure 5-2 we can see the departures and arrivals respectively for June 10th. The ASPM and ASDE-X data compare quite well. The difference per quarter is usually one aircraft, and when for one quarter the ASDE-X count is greater by one, the following quarter is less by one, which means that the difference between the two databases may be just because of a difference in how they account for a movement during the quarter. We also notice occasions where the ASDE-X or ASPM counts exceed the declared ADR or AAR. For a single flight or two every now and then, this is not problematic – it may indicate differences in transcription times for the different methods of data collection, or the fact that stochastic effects play an important role in determining what actually happens at the airport, compared to what might have been predicted a few hours prior.

In Figure 5-3 and Figure 5-4 we can see respectively the departures and arrivals for July 3rd. For the arrivals we can see that ASDE-X counts are almost consistently higher than the ones reported in ASPM, something worth looking into in more detail in order to figure out the reason for this disparity.

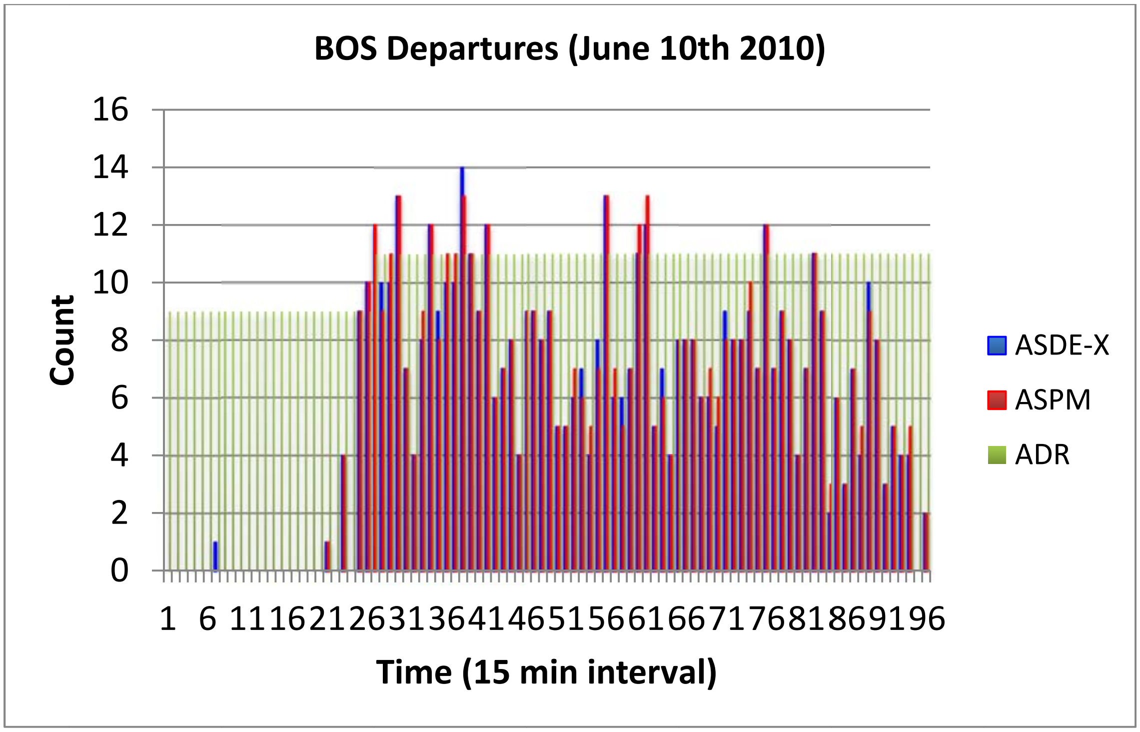

Figure 5-1: ASPM vs ASDE-X comparison for BOS Departures June 10th 2010

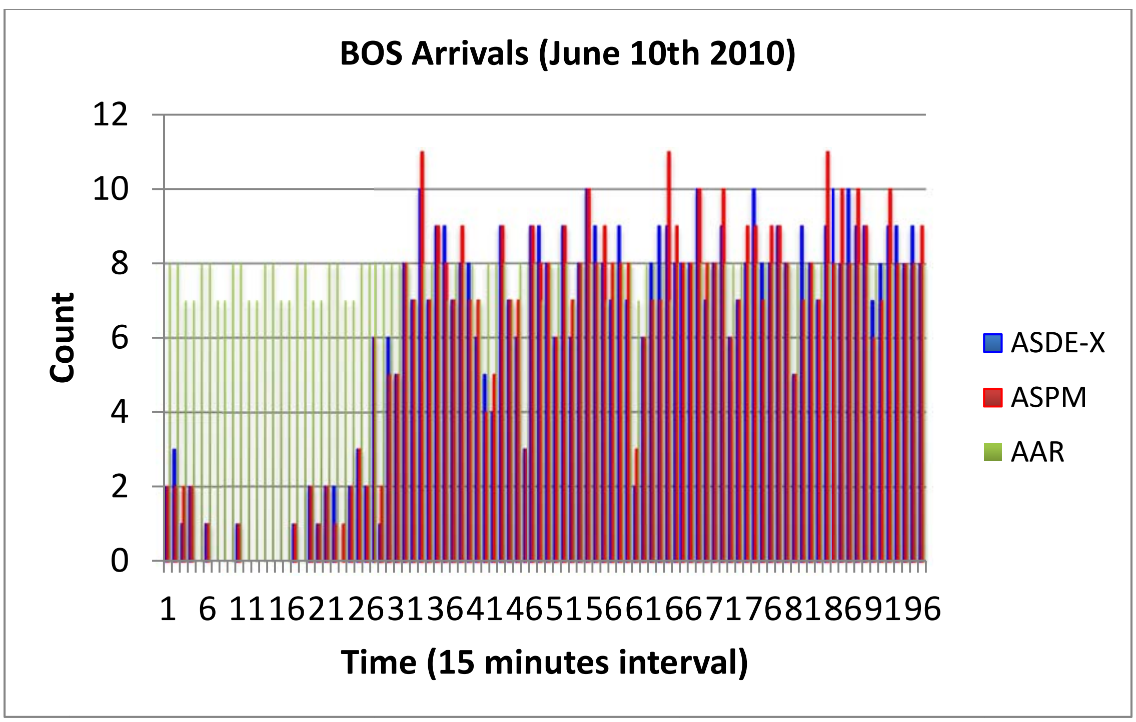

Figure 5-2: ASPM vs ASDE-X comparison for BOS Arrivals June 10th 2010

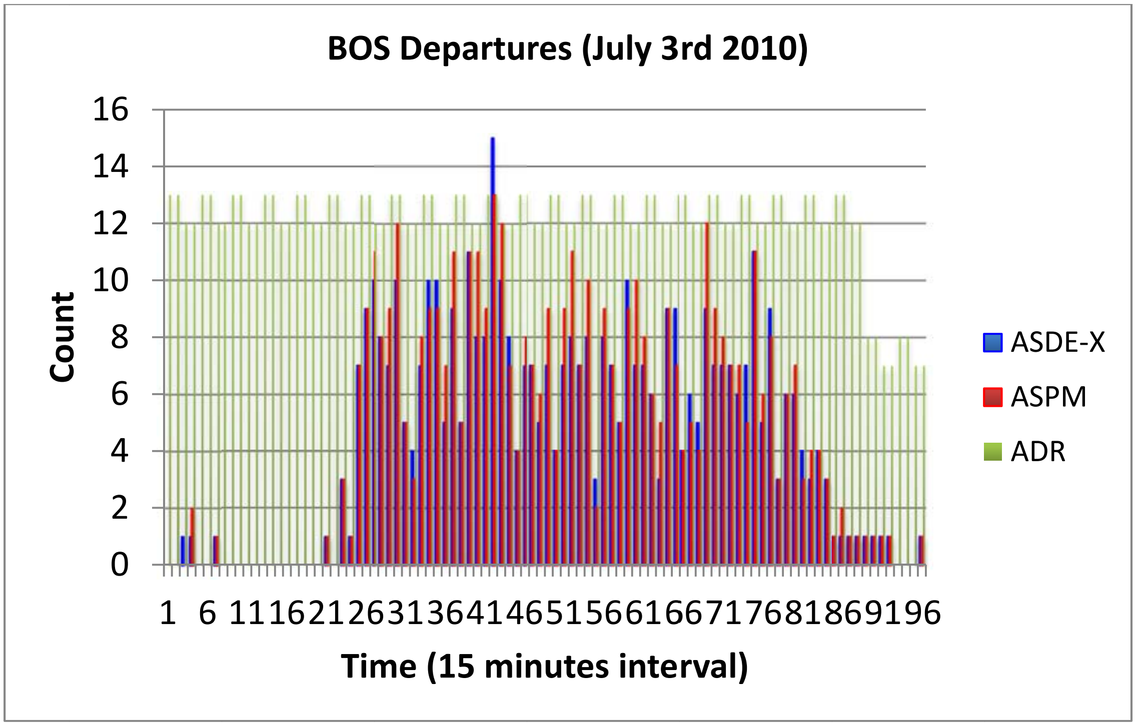

Figure 5-3: ASPM vs ASDE-X comparison for BOS Departures July 3rd 2010

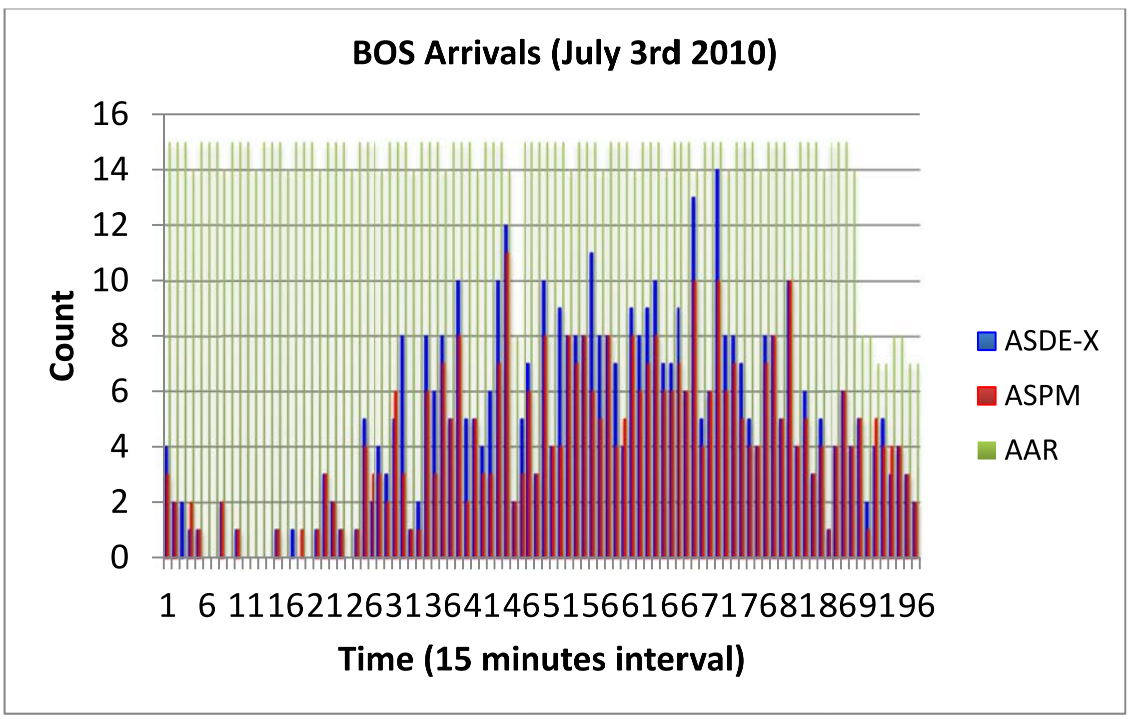

Figure 5-4: ASPM vs ASDE-X comparison for BOS Arrivals July 3rd 2010

In Figures 5-5 through 5-8 we see similar results for July 13th and October 5th. In general the two different databases compare quite well. We see quite a few instances where the actual airport operations never reached the declared capacities, even during the busiest periods of the hour, which suggests that working capacities were in place through some other mechanism (such as a GDP) that were never accurately reflected in the recorded ADR or AAR.

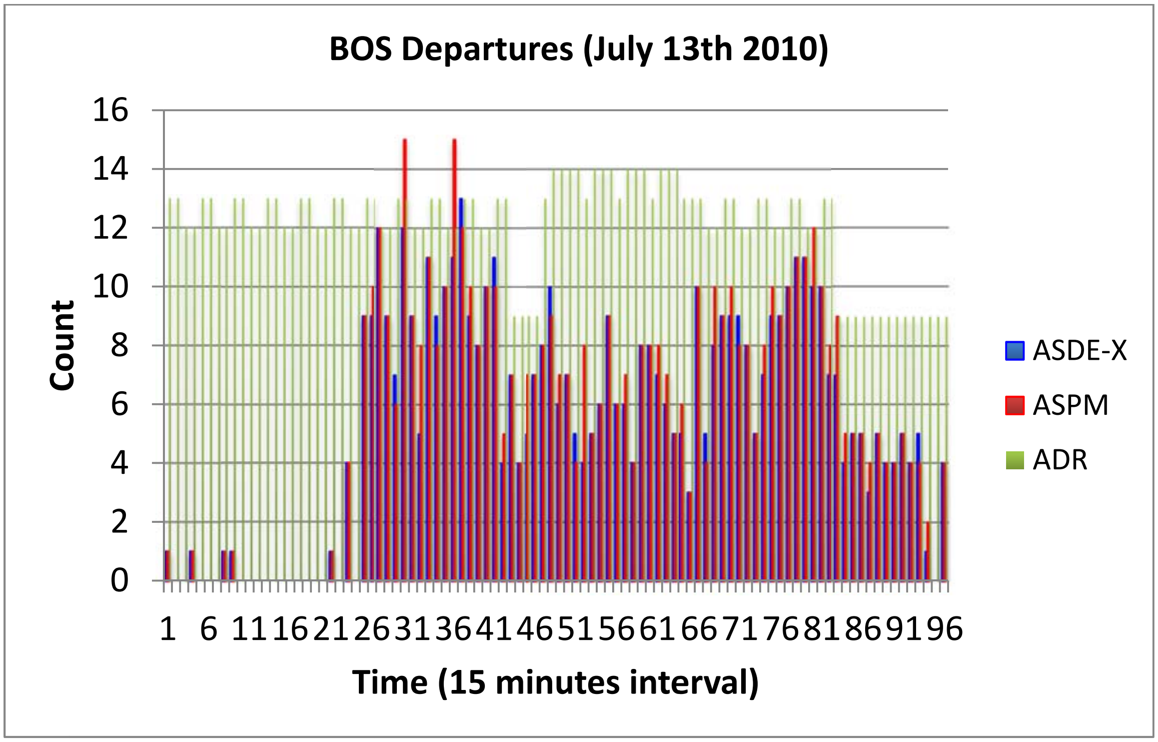

Figure 5-5: ASPM vs ASDE-X comparison for BOS Departures July 13th 2010

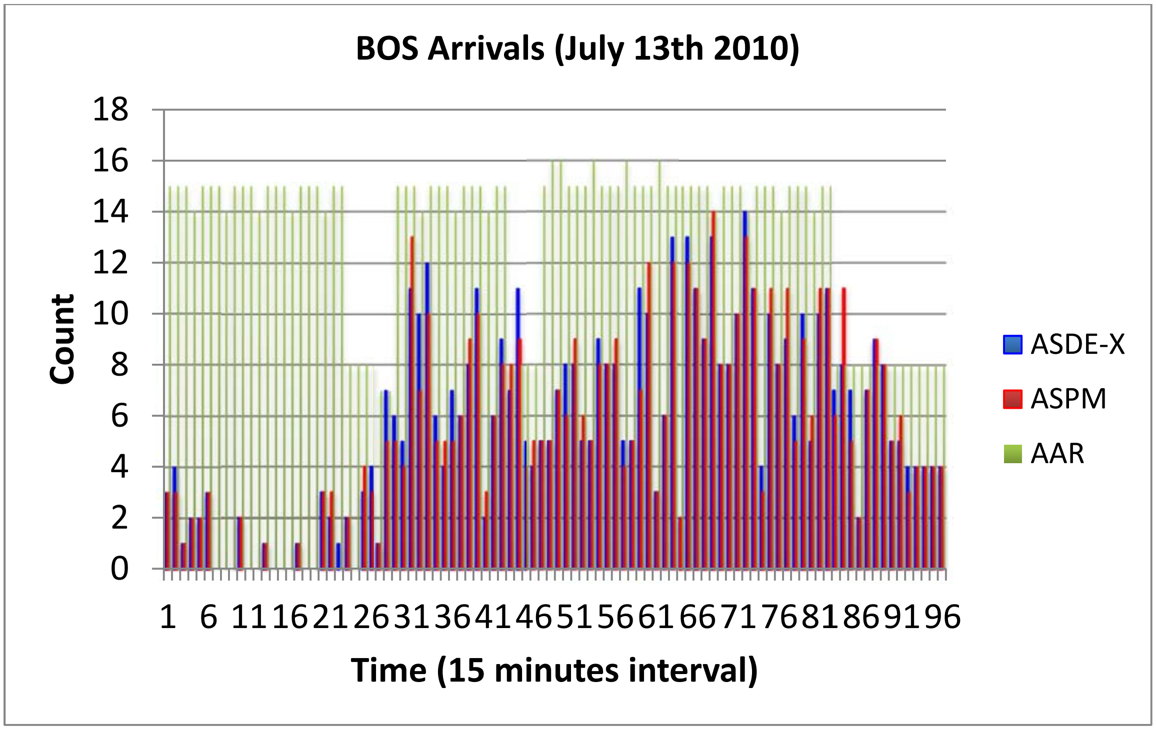

Figure 5-6: ASPM vs ASDE-X comparison for BOS Arrivals July 13th 2010

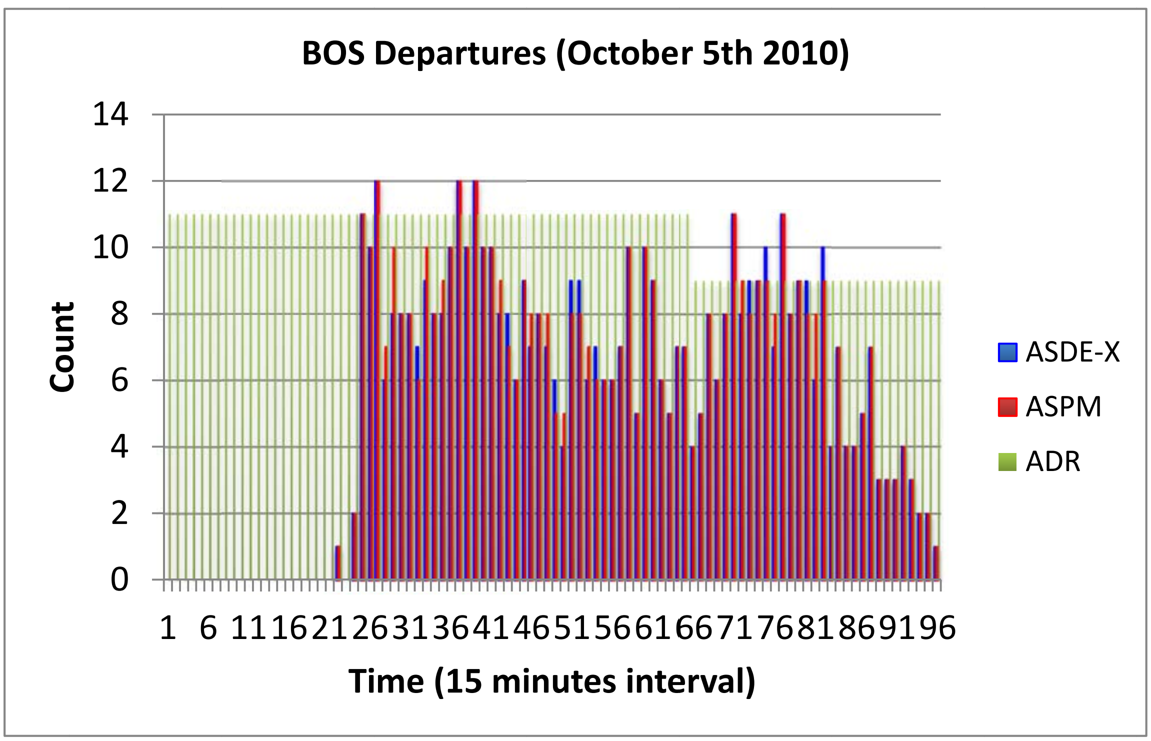

Figure 5-7: ASPM vs ASDE-X comparison for BOS Departures October 5th 2010

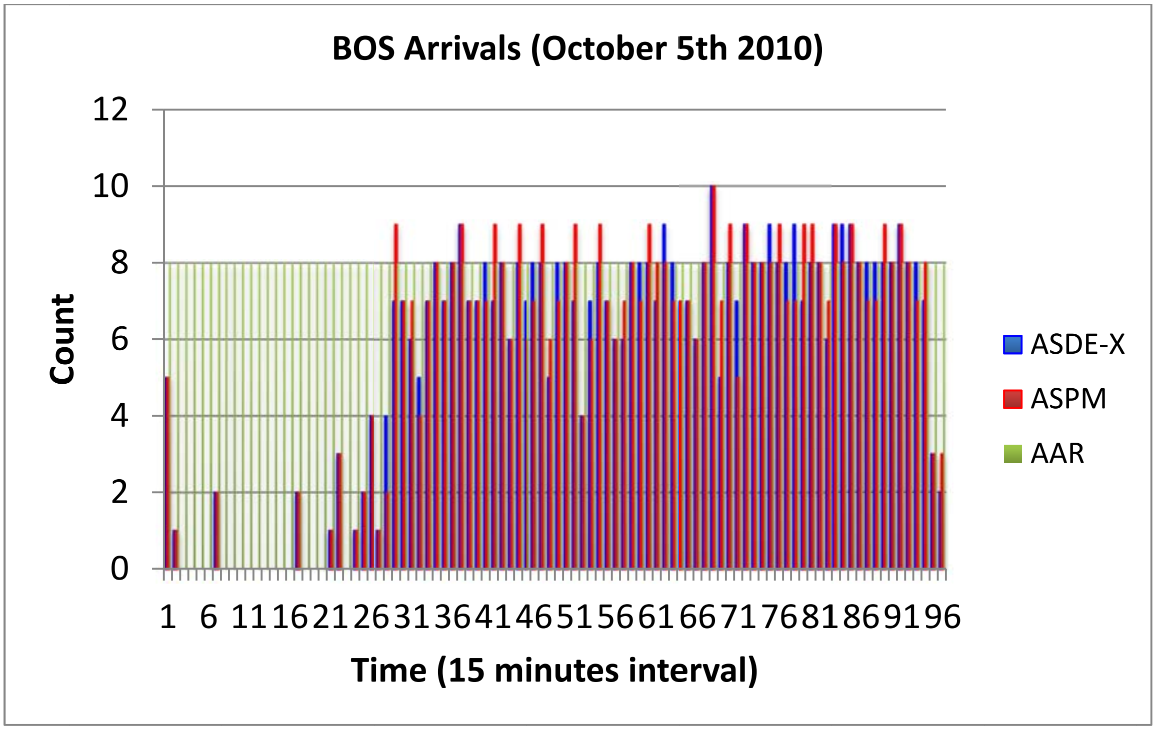

Figure 5-8: ASPM vs ASDE-X comparison for BOS Arrivals October 5th 2010

5.2 Distributions of AARs, ADRs, and ARRs+ADRs

From three years (2009-2011) worth of ASPM data for BOS one can see that numerous runway configurations were used under different meteorological conditions. In Figure 5-9 we can see the layout of the runways in BOS. Some runway configurations were dominant for many quarter hours during that three-year period -(22L, 27|22L, 22R), (22L|22R), (27|22R), (4L,4R|4L, 4R, 9), (4R|4R, 9)- and some others had fewer occurrences -(4R|4L, 9), (4L, 4R, 15R|4L, 4R, 9), (33L|4R), (15R, 15L|9)- .

For a couple of these runway configurations we did some analysis to see the distributions of ARRs, ADRs and the sum of these capacities combined under the same meteorological conditions. As defined in the ASPM database, ADR is the number of departures an airport can support per unit of time. For AAR we used what is called in ASPM the Capacity AAR, which is defined as the Capacity Airport Arrival Rate, or the number of arrivals an airport can support per unit of time. The Capacity AAR is supposed to be used when no traffic management initiatives are in effect. For each runway configuration we broke down the data depending on the meteorological conditions to VMC and IMC and then to VFR, Marginal VFR, Low IFR, and IFR.

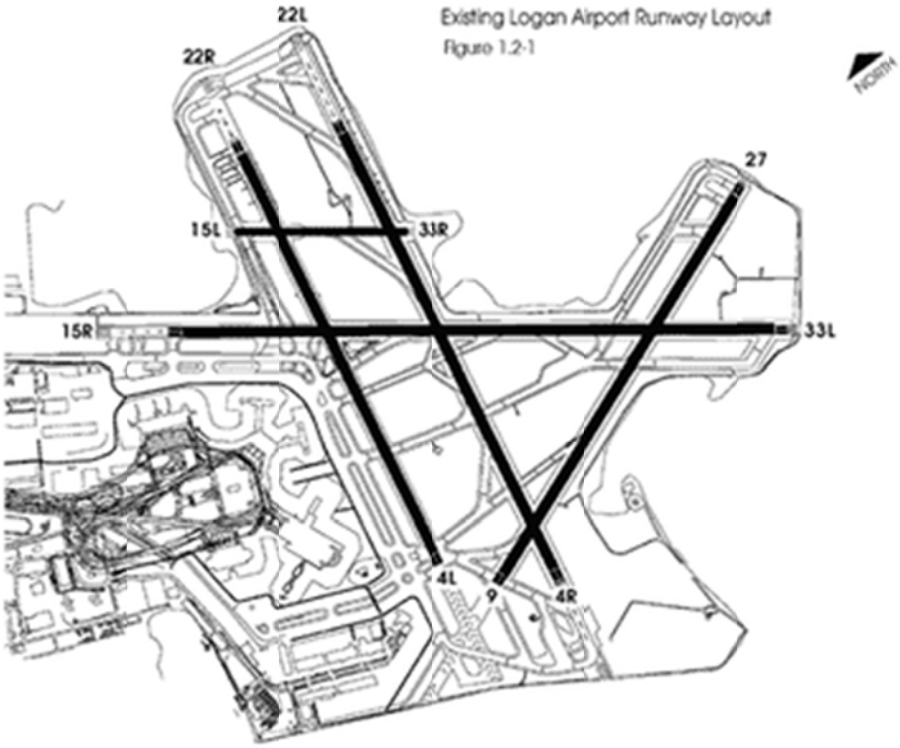

Figure 5-9: Runway configuration in BOS

5.2.1 Arrivals in Runways 15L and 15R, Departures in Runway 9

The first runway configuration we examined is for arrivals using runways 15L and 15R and departures using runway 9. This runway configuration was observed 110 times (quarters) in the three-year period and was used only during VMC and VFR conditions. In Figure 5-10 we can see that despite the same meteorological conditions around BOS, the ADRs declared varied between 7 to 14 aircraft per hour (ac/hr). Most of the times the ADR was set to 9 ac/hr. For ARRs, as we can see in Figure 5-11, the variation is smaller (7 to 11), and most of the time was either 7 or 8 ac/hr. The combined ADRs and ARRs are showed in Figure 5-12 and as we can see this varied between 14 to 23 ac/hr.

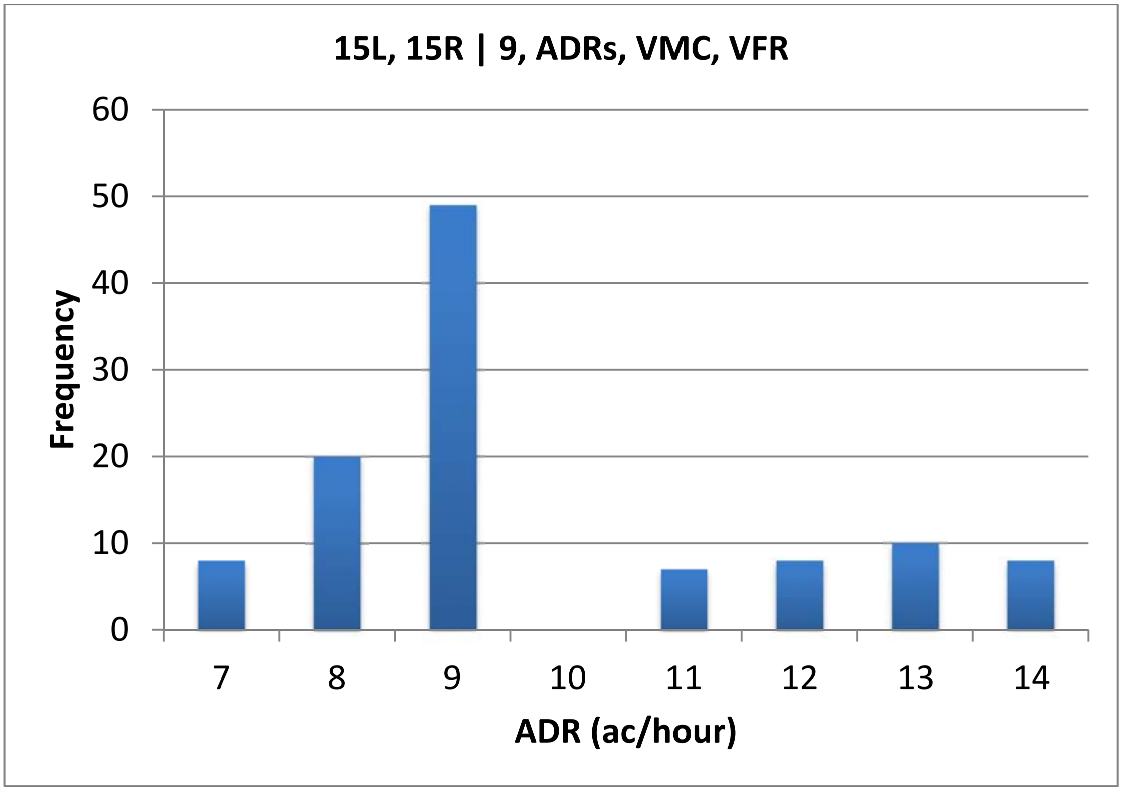

Figure 5-10: ADRs for 15L, 15R|9

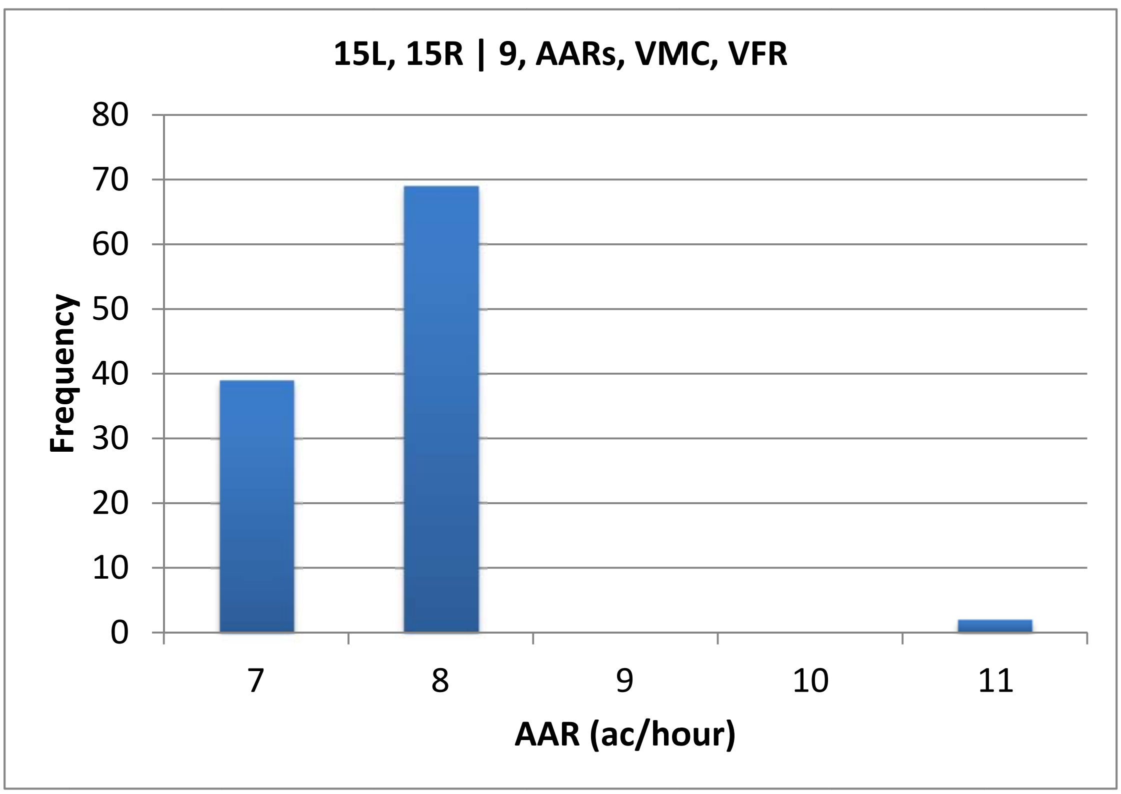

Figure 5-11: ARRs for 15L, 15R|9

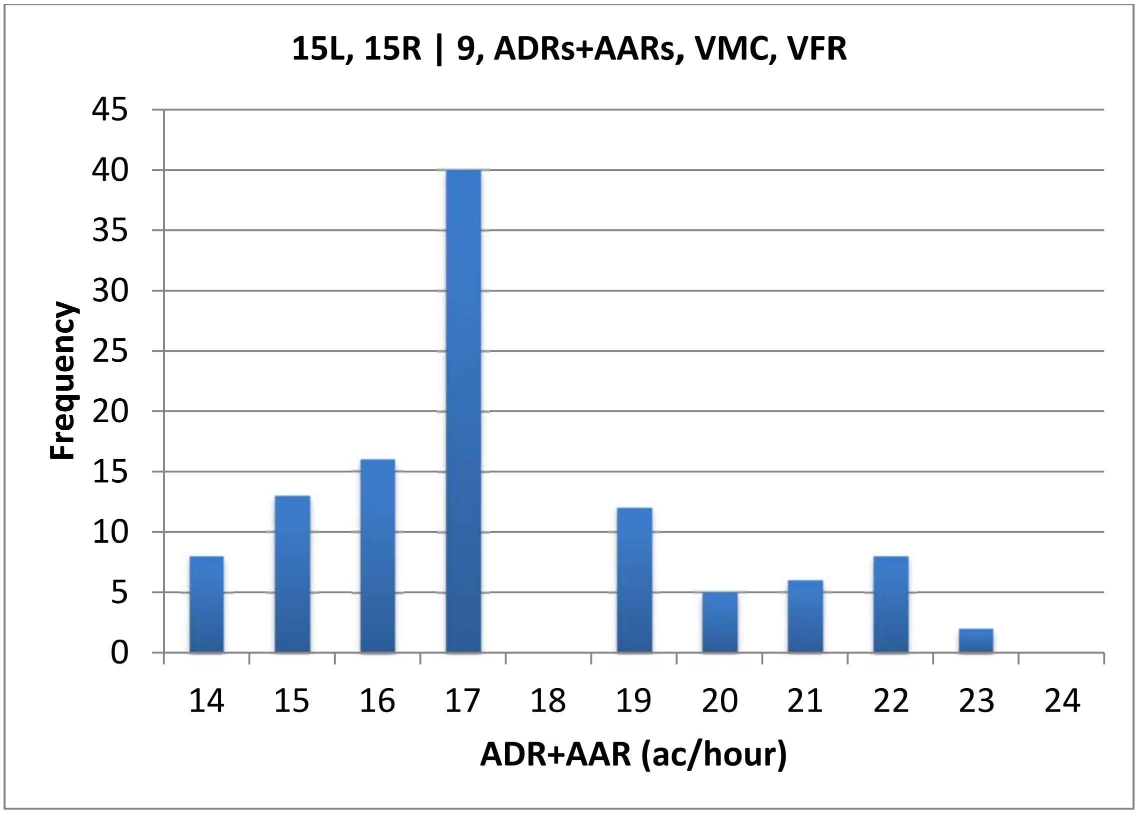

Figure 5-12: ADRs+ARRs for 15L, 15R|9

5.2.2 Arrivals in Runway 22L and Departures in Runway 22R

The next runway configuration we examined was for arrival using runway 22L and departures using runway 22R. During the three-year period this runway configuration was observed 3672 times for VMC weather conditions. Depending on the ceiling and visibility we broke down the observations to VFR and Marginal VFR and looked at the distributions of ADRs, ARRs and both of them combined.

As we can see in Figure 5-13 the declared ADRs varied substantially between 6 and 14 ac/hr for VMC and VFR conditions. Most of the times the ADR was set to 9, 12 or 13. The capacity ARR for this runway configuration during VMC and VFR conditions varied between 6 and 13 ac/hr and the rate that mostly was declared was 8 ac/hr (Figure 5-14). In Figure 5-15 we can see the combined rates, which range between 12 and 27 ac/hr.

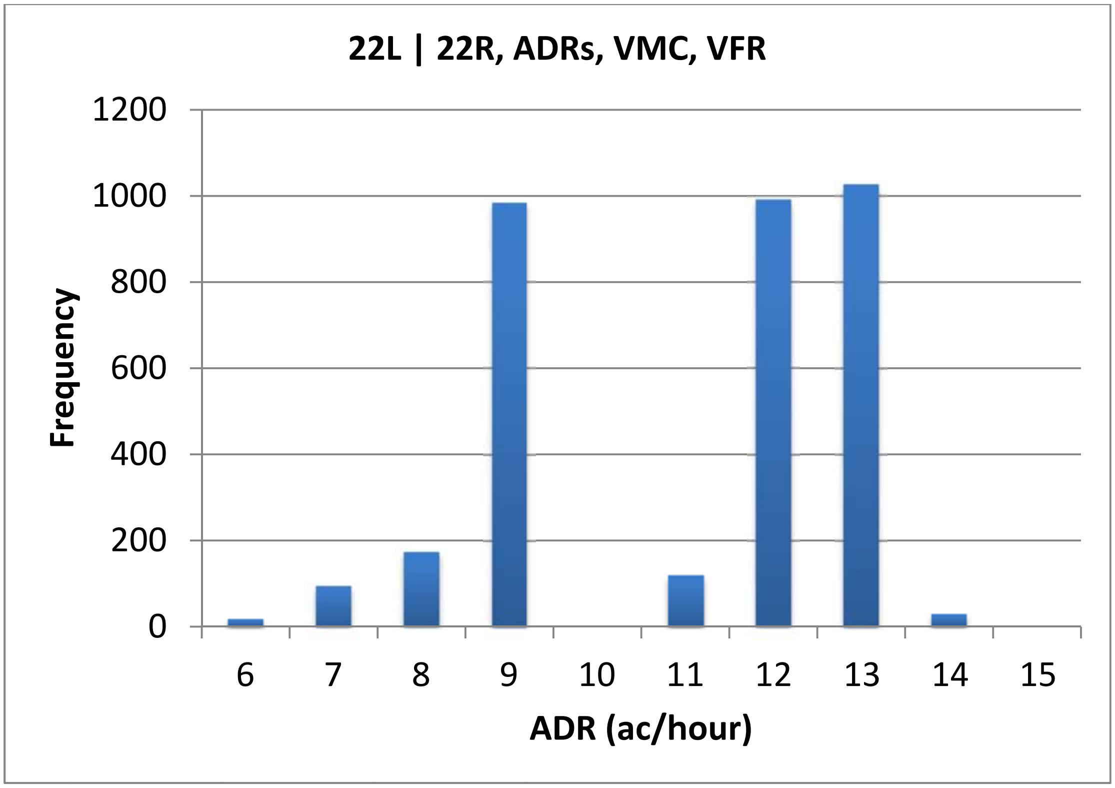

Figure 5-13: ADRs for 22L|22R under VFR conditions

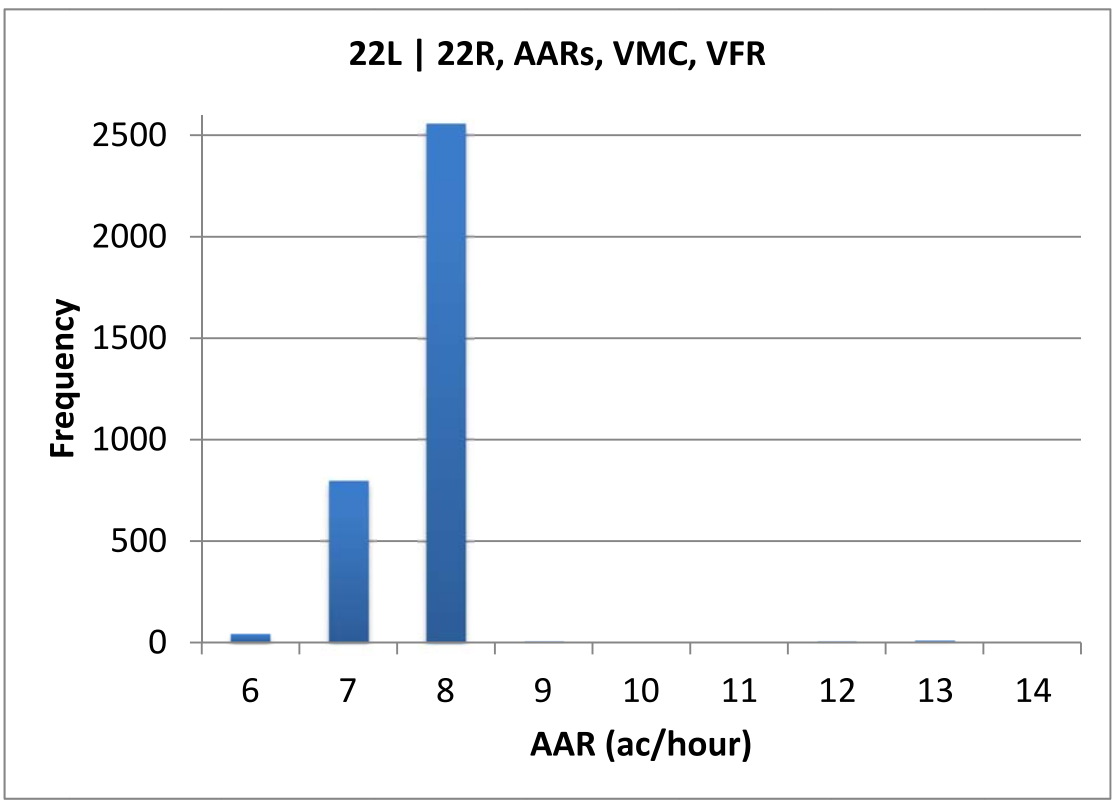

Figure 5-14: ARRs for 22L|22R under VFR conditions

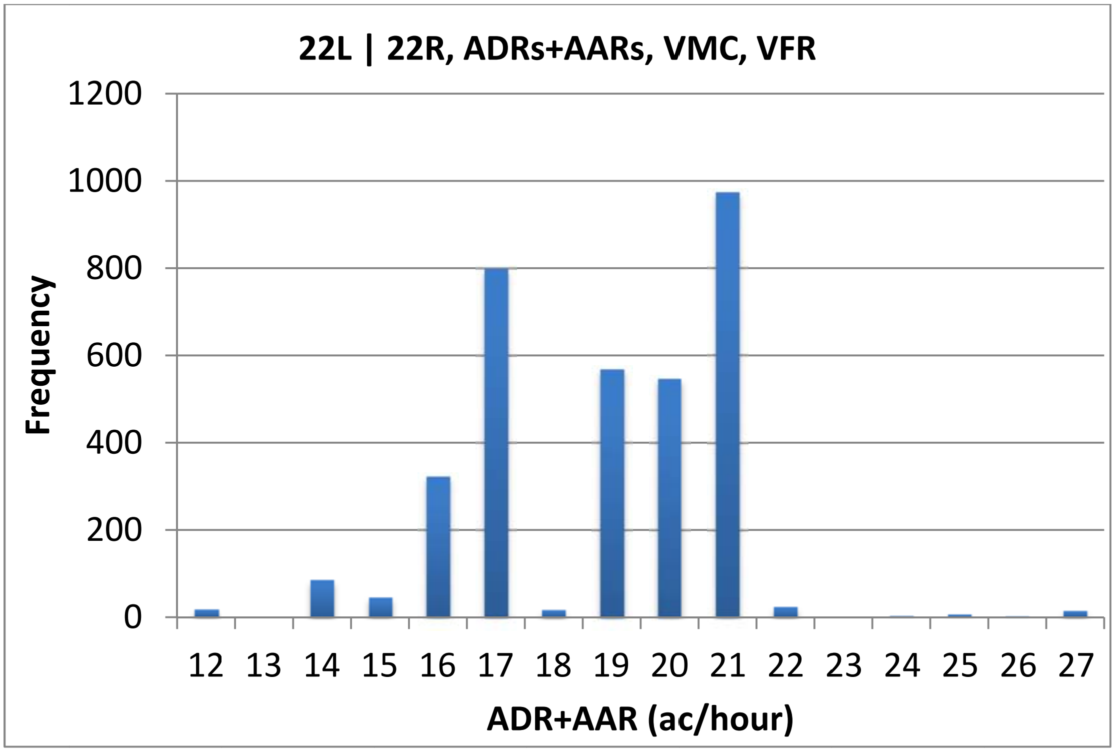

Figure 5-15: ADRs+ARRs for 22L|22R under VFR conditions

The rest of the observations were for VMC and Marginal VFR conditions. In Figure 5-16 we can see the ADRs, which range between 7 and 13. ARRs were either 7 or 8 ac/hr, which is very different from the previous VMC-VFR case, where the ARRs varied much more (Figure 5-17). Finally the combination of ADRs and ARRs are shown in Figure 5-18 and take values that range between 14 and 21 ac/hr.

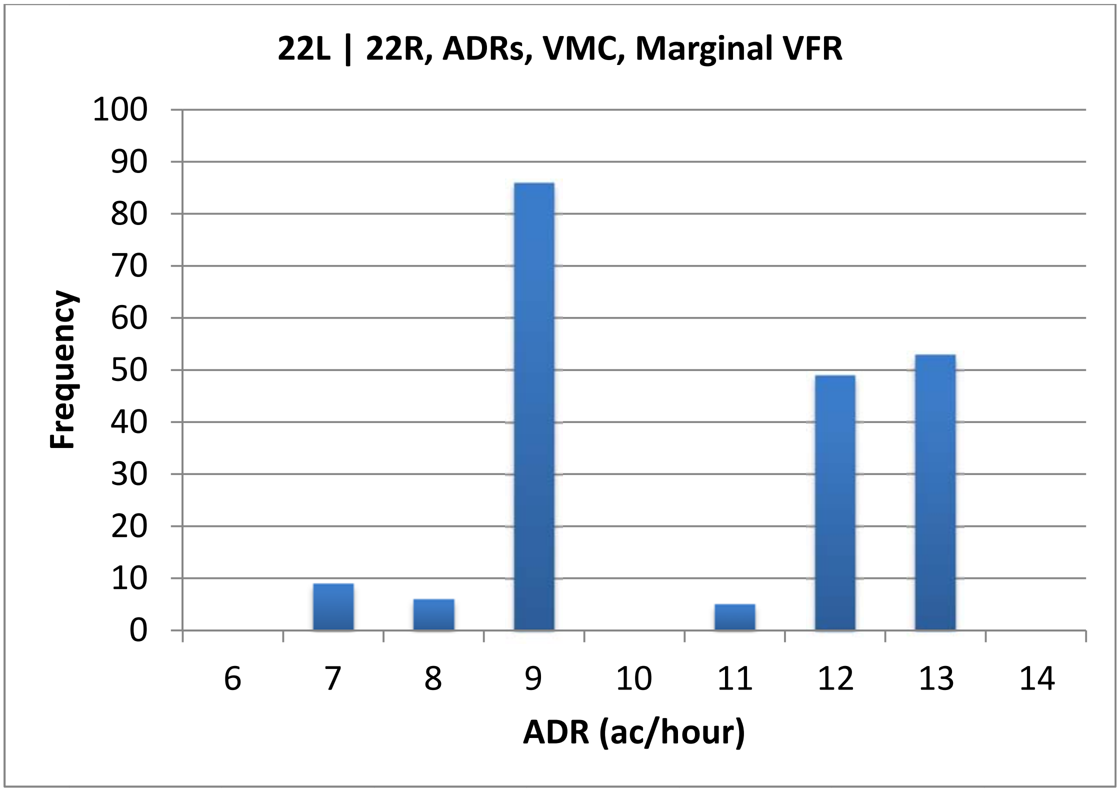

Figure 5-16: ADRs for 22L|22R under Marginal VFR conditions

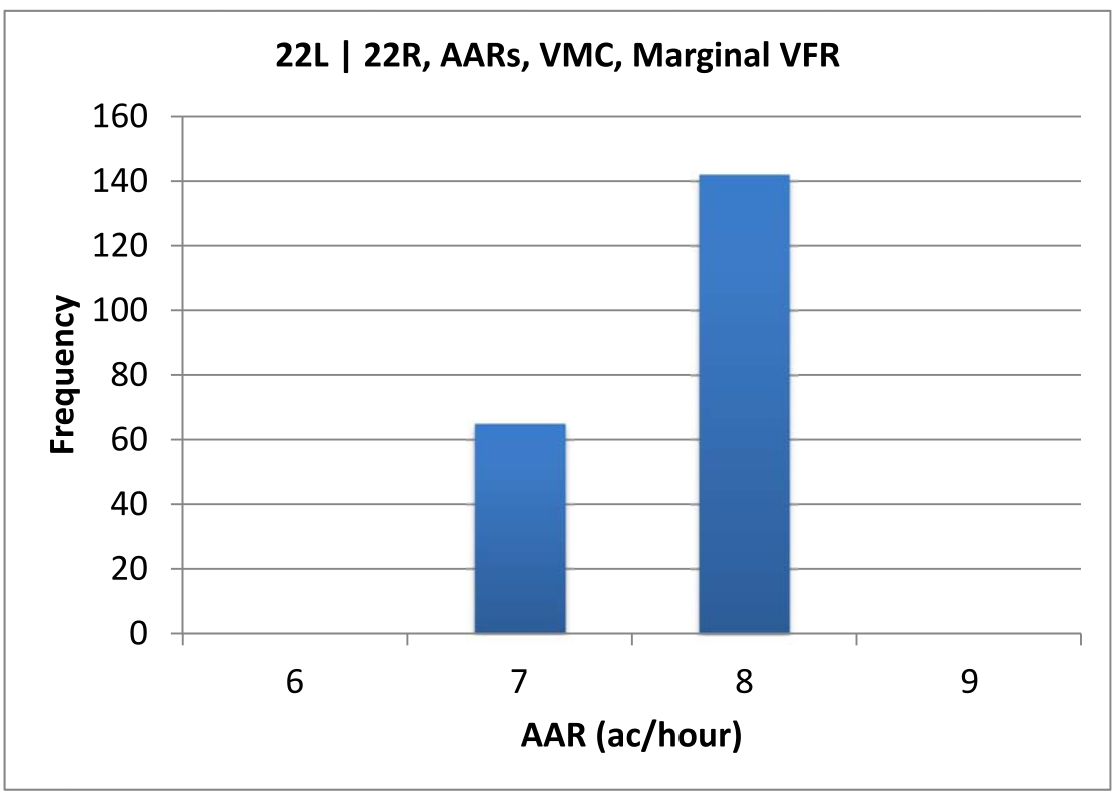

Figure 5-17: ARRs for 22L|22R under Marginal VFR conditions

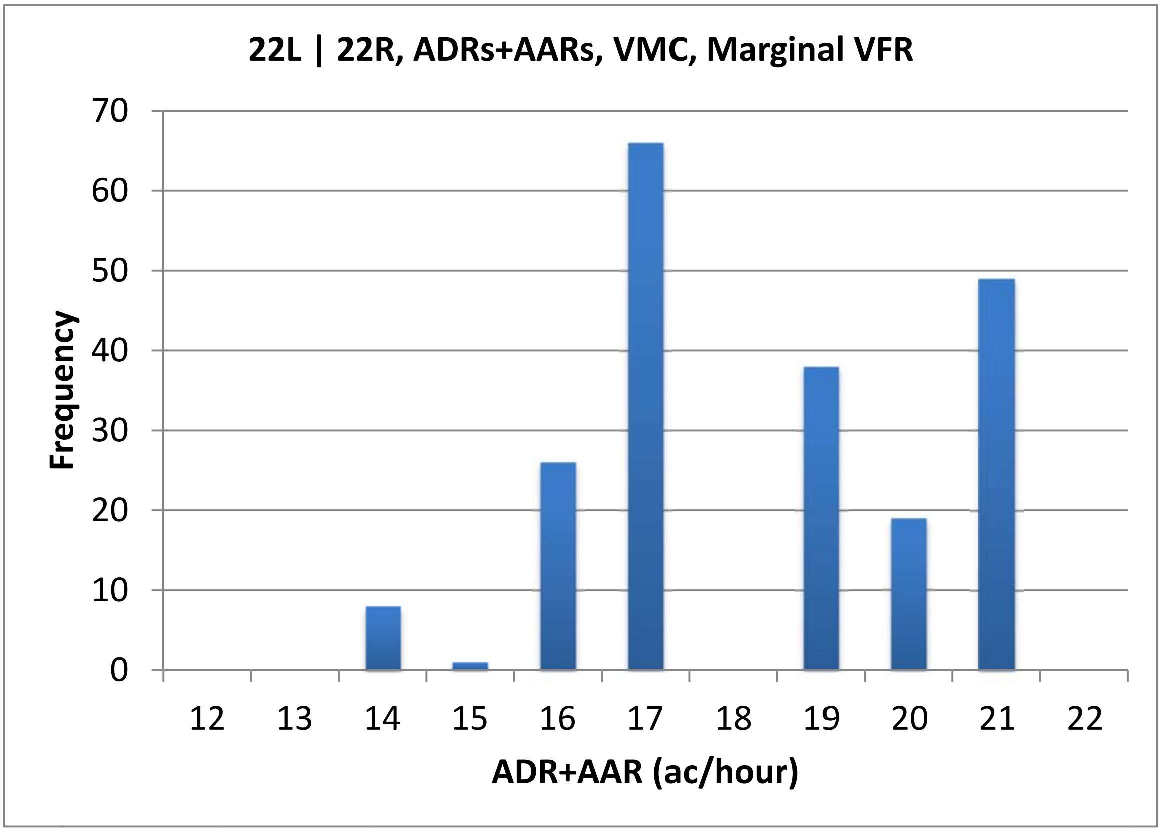

Figure 5-18: ADRs+ARRs for 22L|22R under Marginal VFR conditions

Collectively, the lesson from these figures is that even on the best of days, it is hard to relate declared capacities to what actually happened, even during periods of the day when you would expect sufficient demand to drive the airport to a capacitated state. To be fair, we only show results for a single airport, and a more exhaustive look at this subject would have to consider a much greater variety of facilities and dates.

The subject of most interest, however, was in the rates called for capacities for GDP planning purposes. In particular, we were interested in how these rates might change over the course of an individual GDP, in which several revisions might be possible. Carriers make decisions about swapping and cancelling flights considering the projected delays (which are determined by the projected capacities), plus their own estimates about the reliability of these estimates. To the extent that these parameters could be made more predictable, carriers might make different decisions, presumably to improve the experience of the passengers, as well as to extract economic benefits for themselves. The following subsection addresses our preliminary data analysis along these lines.

5.3 Call Rates and GDP Advisories

Using a custom-written Matlab script, we were able to poll the FAA Ground Delay Program advisory data for GDP initiation, revision, and termination data for any airport, any date. We focused on Newark and San Francisco, as these are two of the airports where GDPs are most frequent. We parsed the data and constructed GDP "trajectories" that essentially reflect the time-dependent projections of GDP rates. Thus, there are two time dimensions in play: any GDP unfolds over time, so even its initial declaration includes a vector of hourly capacities. Over time, however, revisions are issued, which can be thought of as replacing some of the latter portion of that vector with improved estimates, and in many cases, extending the expected duration of the GDP, although shortening is also possible. Table 5-1 shows an example of a GDP evolving over time.

Table 3: Example GDP evolution

Program Period	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**0**	**1**	**2**	**3**	**4**	**5**
Time														
16	**32**	32	32	32	32	32	32	32	32	32	32	32	32	32
17			32	32	32	32	32	32	32	32	32	32	32	32
18														
19														
20														
21														
22							36	36	36	36	36	36	36	36
23														
0									38	38	38	38		

This table reflects a GDP that occurred at Newark (EWR) on December 10, 2012. The GDP was initially declared just before the 16:00 (GMT) hour, which called rates of 32 arrivals per hour for the subsequent 14 hours. In the 17:00 hour, the start time was pushed back to the 18:00 hour. For the next several hours, presumably the GDP proceeded as declared. At the 22:00 hour, however, the called rates were upgraded to 36 for that hour and the remaining 7 hours as well. At the 0:00 hour, the rates were increased again, and the termination time of the GDP was moved two hours earlier. This evolution of called rates shows very directly the kind of unpredictability that a carrier could anticipate when confronted with a GDP. The rates that actually prevailed (32 from 16:00 to 21:00, 36 from 22:00 to 23:00, and 38 from 0:00 to 03:00) were quite different than what was predicted at 16:00. Perhaps the planned rates of 32 were low enough that subsequent delay estimates were severe enough that some carriers cancelled flights, only to learn later that capacities were to improve, albeit too late to reconsider those decisions.

While this table only shows an example of what can happen to a single GDP, Figure 5-19 offers a different look, by illustrating an important artifact of numerous GDPs simultaneously. In this figure, the vertical axis represents the difference between the rate that was predicted, and the number of arrivals that actually used the airport. Each line is a different GDP. The predicted rates are all taken from the initial GDP declaration, which is the set of data that carriers would initially be working with when deciding if they should cancel any flights. The figure includes both positive and negative numbers. Negative numbers imply that more flights arrived than were programmed for; this might be the result of travel time differences, pop-ups, or exempted flights. More alarming are the positive numbers, which indicate that fewer aircraft used the (presumably) congested airport than the reduced capacity afforded by the GDP. It might be the case that certain hours did not have enough demand to fill even the reduced capacity (this would be expected towards the end of the GDP), and perhaps some flights were cancelled with no compression being possible to fill the gaps. It is worrying, however, that this kind of figure might be indicative of important inefficiencies during the most important periods of irregular operations. From a predictability perspective, one might have hoped for a certain convergence of these lines over the course of the GDP evolution, but that is not evidenced either in this figure or in the numerous others for other airports and time periods that we constructed. Of course, this figure does not include any GDP revisions that might manifest that conversion exactly, so that is what we turn to next.

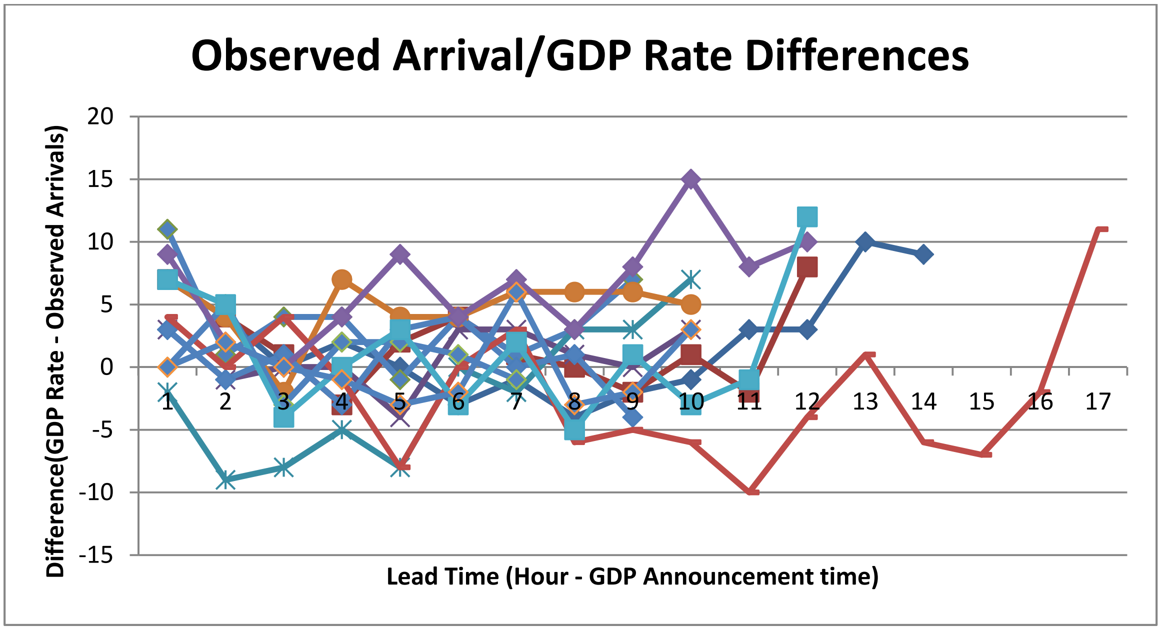

Figure 5-19: GDP rate differentials, EWR airport, January 2010

Figure 5-20 is an attempt to show how accuracy of GDP called rates change over the course of issued GDP revisions. These data are also from the same period of time at EWR. In this case, we do not connect individual GDP trajectories by line, but rather use different series to show the effects of different numbers of GDP revisions over time. The vertical spread for each series, then, is indicative of the uncertainty in predicted rates for that given class of revision. The horizontal axis represents the look-ahead time. If the process of revising GDPs actually produces better capacity estimates, then one would expect the vertical variation to reduce when moving from the initial GDP predictions to the 2nd revision, 3rd revision, and so forth. One could argue that such a pattern is evident in this figure, particularly for time periods in the 7-10 hour look ahead range. However, the overall picture is not as compelling as one might have hoped. There were too few GDPs in this data set (and probably overall) for a 4th revision to play an important role in this analysis. Again, one has to be careful about inferring too much from the most distant look-ahead times. These are the times when the GDPs are terminating anyway, and the demand during these periods might have naturally fallen off to considerably below even the GDP reduced capacities, thereby obviating the GDP itself.

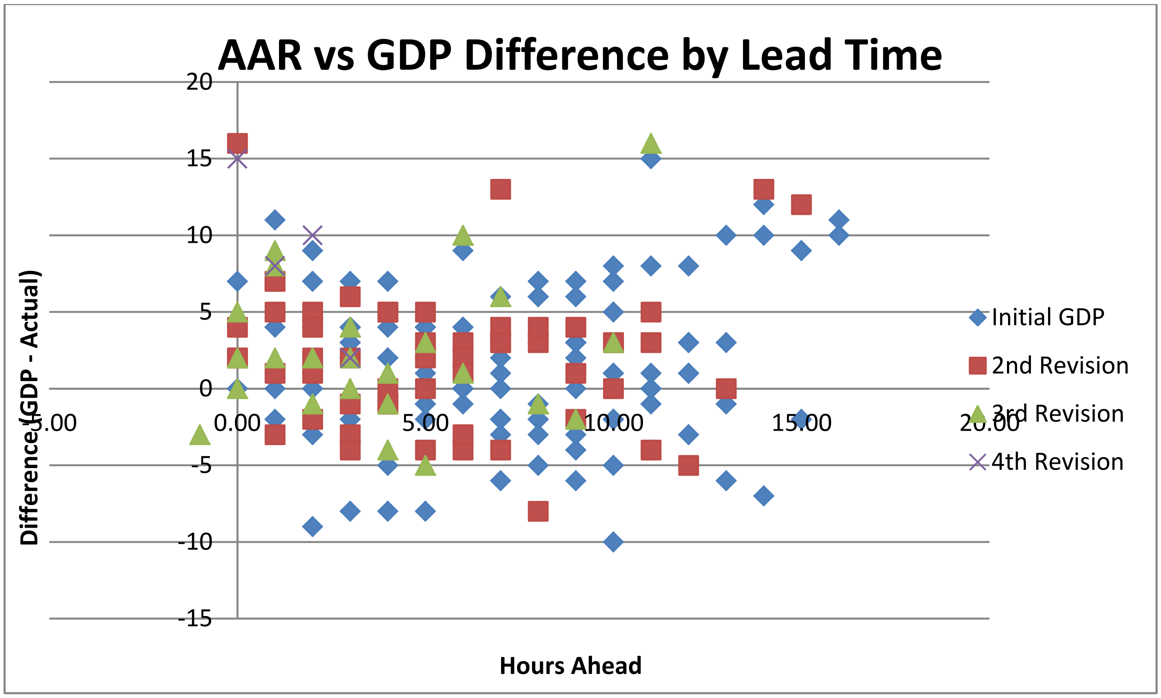

Figure 5-20: GDP rate variability by look-ahead time, EWR airport, January 2010

Ultimately, this is the extent to which this analysis proceeded. We have identified what we think are some useful ways to parse the data, and some interesting ways to plot it graphically. The number of dimensions in the problems exceeds the number that can be shown in pictures, however, so it is complicated to find the picture that captures the most salient aspects of the problem most usefully. We have also illuminated some of the sources of uncertainty that will likely cloud any such analysis.

A logical next step would be to do a deep dive into similar data for a wide variety of airports and time periods, supported by scripts to automate the processes of parsing the data and plotting the results. There may still be important patterns hidden in the data that could be mined and revealed with such an effort.

6 Conclusions

In this research we have investigated predictability as an aspect of NAS performance. More specifically, we have examined two benefit mechanisms by which improved predictability may lead to reduced cost to airlines. One mechanism is based upon airline scheduled block time setting, and the other on airline fuel loading practice. We have considered both the behavioral components of these mechanisms—for example how flight time variability affects scheduled block times—and the economic components—for example how scheduled block time influences airline operating cost.

In both instances, sizable benefits from improving predictability are found. In the case of scheduled block time setting, an annual saving of roughly \$150 million to United, American and Delta could be saved from a plausible change in the distribution of realized block times. With fuel loading, we estimate that Delta could save \$30-\$40 million per year in the form of reduced fuel burn if flight time variability could be eliminated.

Extrapolation of either of these figures to the entire industry is hazardous. Using available seat miles as a basis, in 2010 US airlines together produced 2.5 times the output as United, American, and six times the output of Delta itself. Applying these multipliers yields \$400 million in savings for reductions in scheduled block time and \$67-252 million in savings from reduced fuel burn. However, it should be emphasized that there is a wide variety of cost structures and fuel loading practices across the industry, making these estimates very rough.

Be this as it may, the magnitudes of these estimates shy in comparison to the total cost of delay to US carriers, estimated to be on the order of \$10 billion (Zou and Hansen, 2012). While it would be a mistake to ignore predictability as an aspect of NAS performance, it would also be a mistake to make too much of it.

This report has not directly considered the question of metrics. Nonetheless, our results have an important bearing on this matter. The most important finding is that the proper metric depends on which benefit mechanism is being considered. In the case of scheduled block time setting, the most relevant metrics concern the inner right tail of the flight time distribution. An example metric would be the difference between the median and 70th percentile of the realized block time. While the region of the distribution to the right of the 70th percentile is also important, it is mainly as a driver of delays against schedule rather than the schedule itself. For the same reason, the standard deviation of the flight time is not a very good metric when it comes to the scheduled block time effect, because the variance is heavily influenced by the far right tail.

By the same token, standard deviation is an appropriate metric for fuel loading. For obvious reasons, airlines cannot afford to ignore the far right tail of the distribution in this context. Indeed, it may be that a metric focused on the far right tail (rather than just strongly influenced by it, like the standard deviation) would be a better choice. This might be a topic for future research.

This report has focused on strategic predictability, reflecting variability in flight times over periods of months or years. For the most part, we have not considered tactical predictability, which reflects the ability to predict flight times and delays on the day of operations. One could argue that the fuel savings component is both, as the decisions are made on the day of operations, but ultimately they result from the uniform application of a corporate strategy that is decided well in advance. Strictly tactical predictability is difficult to measure since it requires knowledge of what flight operators knew and when they knew it. It is also hard to monetize. Research in both of these areas might be considered in the future. We have included a short chapter reflecting some preliminary data analysis and data visualization aimed at predictability of called rates for Ground Delay Programs. This line of inquiry is only in its infancy, but we believe this is a rich area for future research.

7 References

Abdelghany, K., A. Abdelghany, and S. Raina, 2005. A model for airline fuel management strategies. *Journal of Air Transport Management*, 11, 199-206.

Abrantes, P. and M. Wardman, 2011. Meta-analysis of UK values of travel time: An update. *Transportation Research Part A: Policy and Practice,* 45(1), 1-17.

Airlines for America, Annual Crude Oil and Jet Fuel Prices. http://www.airlines.org/Pages/Annual-Crude-Oil-and-Jet-Fuel-Prices.aspx. Last accessed: July 2013.

Aviation Daily, 2013. Aviation week intelligence network market briefing. Aviation Daily, McGraw-Hill Financial, Inc.

Bates, J., J. Polak, P. Jones, and A. J. Cook, 2001. The valuation of reliability for personal travel. *Transportation Research Part E: Logistics and Transportation Review*, 37(2), 191-229.

Bolczak, C. N., J. H. Hoffman, A. J. Jensen, and W. W. Trigeiro, 1997. National Airspace System Performance Measurement: Overview. MITRE Technical Report W0000035, The MITRE Corporation, McLean, VA.

Börjesson, M. and J. Eliasson, 2011. On the use of "average delay" as a measure of train reliability. *Transportation Research Part A: Policy and Practice*, 45(3), 171-184.

Börjesson, M., J. Eliasson, and J. P. Franklin, 2012. Valuations of travel time variability in scheduling versus mean-variance models. *Transportation Research Part B: Methodology*, 46(7), 855-873.

Borokhov, P., S. Blandin, S. Samaranayake, O. Goldschmidt, and A. Bayen, 2011. An adaptive routing system for location-aware mobile devices on the road network. 14th IEEE Conference on Intelligent Transportation Systems, Washington, D.C.

Brownstone, D. and K. Small, 2005. Valuing time and reliability: Assessing the evidence from road pricing demonstrations. *Transportation Research Part A: Policy and Practice*, 39(4), 279-293.

Bureau of Transportation Statistics (BTS). http://www.transtats.bts.gov/

Cao, Y., T. Kotegawa, D. Sun, D. DeLaurentis, and J. Post, 2011. Evaluation of Continuous Descent Approach as a standard terminal airspace operation.Ninth USA/Europe Air Traffic Management Research and Development Seminar, Berlin, Germany.

Carrion, C. and D. Levinson, 2012. Value of travel time reliability: A review of current evidence. *Transportation Research Part A*, 46(4), 720-741.

Chiraphadhanakul, V. and C. Barnhart, 2013. Robust flight schedules through slack re-allocation. *EURO Journal on Transportation and Logistics*, 2(4), 277-306 .

Ciuperca, G. and V. Girardin, 2005. On the estimation of the entropy rate of finite Markov chains, Proceedings of the International Symposium on Applied Stochastic Models and Data Analysis, Brest, France.

Clarke, J. P. B., N. T. Ho, L. Ren, J. A. Brown, K. R. Elmer, K.-O. Tong, and J. K. Wat, 2004. Continuous Descent Approach: Design and flight test for Louisville International Airport. *Journal of Aircraft*, 41(5), 1054-1066.

Cohen, H. and F. Southworth, 1999. On the measurement and valuation of travel time variability due to incidents on freeways. *Journal of Transportation Statistics*, 2(2), 123 – 131.

Cover, T., and J. Thomas, 1991. Elements of Information Theory. John Wiley and Sons, Inc., New York, NY.

Coy, S., 2006. A global model for estimating the block time of commercial passenger aircraft. *Journal of Air Transport Management*, 12(6), 300-305

Deshpande, V. and M. Arikan, 2012. The impact of airline flight schedules on flight delays. *Manufacturing and Service Operations Management*, 14(3), 423-440.

U.S. National Archives and Records Administration, 2012. Code of Federal Regulations, Title 14. Aeronautics and Space, section 1.1.

EPA, 2013a. Emission Factors for Greenhouse Gas Inventories, US Environmental Protection Agency. http://www.epa.gov/climateleadership/documents/emission-factors.pdf. Last accessed: July 2013.

EPA, 2013b. U.S. Transportation Sector Greenhouse Gas Emissions. EPA, September, 2013 http://www.epa.gov/otaq/climate/documents/420f13033a.pdf

Ettema, D. and H. Timmermans, 2006. Costs of travel time uncertainty and benefits of travel time information: conceptual model and numerical examples. *Transportation Research Part C: Emerging Technologies*, 14, 335 – 350.

FAA, 2012. Destination 2025. Federal Aviation Administration, U.S. Department of Transportation, Washington, D.C.

Fosgerau, M. and A. Karlstrom, 2010. The value of reliability. *Transportation Research Part B: Methodology*, 44(1), 38-49.

Gaver, D., 1968. Headstart strategies for combating congestion. *Transportation Science*, 2(2), 172-181.

Gulding, J., D. Knorr, M. Rose, J. Bonn, P. Enaud, and H. Hegendoerfer, 2009. US/Europe comparison of ATM-related operational performance. 8th USA/Europe Air Traffic Management Research and Development Seminar, Napa, California.

Hao, L. and M. Hansen, 2013a. Flight time predictability: Concepts, metrics and impact on scheduled block time. 92nd Annual Meeting of the Transportation Research Board, January 2013, Washington, DC.

Hao, L. and M. Hansen, 2013b. Block time reliability and scheduled block time setting. Working paper, University of California, Berkeley.

Hollander, Y., 2006. Direct versus indirect models for the effects of unreliability. *Transportation Research Part A: Policy and Practice*, 40(9), 699-711.

International Civil Aviation Organization (ICAO), 2005. Report: Global Air Traffic Management Operational Concept, 1st Edition.

Jacquet, P., G. Seroussi, and W. Szqankowski, 2008. On the entropy of a hidden Markov process. *Theoretical Computer Science*, 395, 203 – 219.

Jamin, S., A. Schäfer, M. E. Ben-Akiva, and I. A. Waitz, 2004. Aviation emissions and abatement policies in the United States: A city-pair analysis. *Transportation Research Part D: Transport and Environment*, 9(4), 295-317.

Khadilkar, H. and H. Balakrishnan, 2012. Estimation of aircraft taxi fuel burn using flight data recorder archives. *Transportation Research Part D: Transport and Environment*, 17(7), 532-537.

Kho, S. Y., J. S. Park, Y. H. Kim, and E. H. Kim, 2005. A development of punctuality index for bus operation. *Journal of the Eastern Asia Society for Transportation Studies*, 6, 492 – 504.

Knight, T., 1974. An approach to the evaluation of changes in travel unreliability: A safety margin hypothesis. *Transportation*, 3(4), 393-408.

Kwon, J., B. Coifman, and P. Bickel, 2000. Day-to-day travel-time trends and travel-time prediction from loop-detector data. *Transportation Research Record: Journal of the Transportation Research Board*, 1717, 120 – 129.

Leigh, R. J., 1995. Economic benefits of terminal aerodrome forecasts (TAFs) for Sydney Airport, Australia. *Meteorological Applications*, 2, 239-247.

Liston-Hayes, C. and A. Pilkington, 2004. Inventive concentration: An analysis of fuel cell patents. *Science and Public Policy*, 31(1), 15-25.

Liu, Y. and M. Hansen, in press. Ground Delay Program performance evaluation. *Transportation Research Record: Journal of the Transportation Research Board*.

Lomax, T., D. Schrank, S. Turner, and R. Margiotta, 2003. Selecting Travel Reliability Measures. Texas Transportation Institute, College Station, TX.

Moreno-Hines, Felipe, and Deborah Kirkman (2013). Assessing the Nextgen avionics business case from the airline perspective: The implications of airline responses to changes in operational performance. 10th USA/Europe Air Traffic Management Research and Development Seminar (ATM2013), Chicago, IL, USA.

Nikoleris, T., G. Gupta, and M. Kistler, 2011. Detailed estimation of fuel consumption and emissions during aircraft taxi operations at Dallas/Fort Worth International Airport. *Transportation Research Part D: Transport and Environment*, 16 (4), 302-308.

Noland, R.B. and K. Small, 1995. Travel time uncertainty, departure time and the cost of the morning commute. *Transportation Research Record*, 1493, 150-158.

Noland, R., and J. Polak, 2002. Travel time variability: A review of theoretical and empirical issues. *Transport Reviews*, 22 (1), 39-54.

Pells, S., 1987. The Evaluation of Reductions in Travel Time Variability. Ph.D. Dissertation, University of Leeds, UK.

Rice, J. and E. V. Zwet, 2004. A simple and effective method for predicting travel times on freeways. *IEEE Transactions on Intelligent Transportation Systems*, 5(3), 200 – 207.

Riikka, S. and J. Paavilainen, 2010. Measuring railway traffic punctuality from the passenger's perspective. 12th World Conference in Transportation Research, Lisbon, Portugal.

Ryerson, M. S. and M. Hansen, 2010. The potential of turboprops for reducing aviation fuel consumption. *Transportation Research Part D: Transport and Environment*, 15(6), 305-314.

Ryerson, M.S., M. Hansen, and J. Bonn, 2011. Fuel consumption and operational performance. Ninth USA/Europe Air Traffic Management Research and Development Seminar, Berlin, Germany.

Ryerson, M. S. and A. Churchill, 2013. Aircraft rerouting due to abrupt facility outages: A case study of the 2011 Great Tōhoku Earthquake. *Transportation Research Record: Journal of the Transportation Research Board*, 2336, 27-35.

Shannon, C. E., 1948. A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379-423, 623-656.

Simaiakis, I. and H. Balakrishnan, 2010. Impact of congestion on taxi times, fuel burn, and emissions at major airports. *Transportation Research Record: Journal of the Transportation Research Board*, 2184, 22-30.

Simaiakis, I., M. Sandberg, H. Balakrishnan, and R. J. Hansman, 2012. Design, testing and evaluation of a pushback rate control strategy. Proceedings of the 5th International Conference on Research in Air Transportation (ICRAT), Berkeley, CA.

Small, K., 1982. The scheduling of consumer activities: Work trips. *American Economic Review*, 72(3), 467-479.

Sohoni, M. G., D. Klabjan, and Y.-C. Lee, 2011. Robust airline scheduling under block time uncertainty. *Transportation Science*, 45(4), 451-464.

Taylor, M.A.P., 1982. Travel time variability- the case of two public modes. *Transportation Science* 16(4), 507 – 521.

Vickrey, W., 1969. Congestion theory and transportation investment. *The American Economic Review*, 59(2), 251-260.

Wardman, M., 2001. A review of British evidence on time and service quality valuations. *Transportation Research Part E: Logistics and Transportation Review*, 37, 107-128.

Wojcik, L.A., S. L. Mondoloni, S. J. Agbolosu-Amison, and P. Wang. Flexibility Metrics and Their Application to Departure Queue Management. Tenth USA/Europe Air Traffic Management Research and Development Seminar, Chicago, IL.

Zhang, X. and J. A. Rice, 2003. Short-term travel time prediction. *Transportation Research Part C: Emerging Technologies*, 11, 187 – 210.

Zou, B. and M. Hansen, 2012. Impact of operational performance on air carrier cost structure: Evidence from US airlines. *Transportation Research Part E: Logistics and Transportation Review*, 48(5), 1032-1048.