DETERMINING SEASONAL DELAY CURVES

Al Meilus, Federal Aviation Administration, Office of Performance Analysis, Washington, DC

Abstract

This paper presents an approach to expand on the Annual Service Volume (ASV) method for estimating air traffic delays. The ASV approach is a direct and intuitive method to estimate air traffic delays at an airport based on airline scheduled demand levels at that particular airport. The Seasonal Delay Curve (SDC) approach attempts to build on the ASV delay estimates by accounting for seasonal capacity variations. The result is an equally intuitive method that sheds light on the impact that seasons can have on delays at specific airports. This improvement helps explain why a particular scheduled level of demand can produce different amounts of delay in January than in July. This paper demonstrates how delay estimates can differ by season at airports like SFO and ORD that have noticeable seasonality in their arrival and departure rates.

Introduction

In response to a sudden increase in delays experienced in the summer of 2007, the FAA committed to taking a pro-active approach to monitoring airline scheduling, creating an initiative to identify airports forecasted to have chronic delay in the next six months [1]. The FAA has since extended the effort to project delays up to 12 months in advance. This effort also supports a more recent requirement, Section 413 of the FAA Reform and Modernization Act of 2012 (PL 112-95) enacted on February 14, 2012 [2], which mandates that the Administrator take steps to reduce schedules when operations exceed the maximum departure and arrival rate at an airport. To comply with these requirements, since 2009, the Air Traffic Organization (ATO) has been producing a monthly delay report [3], monitoring scheduled operations and delays at the 30 core airports and noting significant trends over a 12-month look-ahead period. The report uses demand projections to identify airports likely to experience significant increase in delays. This paper describes an enhancement to one of the methods used by the FAA to forecast monthly delays.

Current Methods

The ATO uses two models to forecast monthly delays. The models use a demand forecast based on published airline schedules, a near-term demand forecast provided by FAA's Office of Aviation Policy and Plans (APO), and historical operational data. The resulting demand projections are then combined with expected airport capacities to estimate future delays. The models estimate delays in two ways. The first approach uses a representation of airport capacity, called the Annual Service Volume (ASV) [4]. The second approach uses detailed simulation of activity each day to project delays [5]. This paper discusses a potential extension to the ASV approach by incorporating seasonal variation in airport capacity.

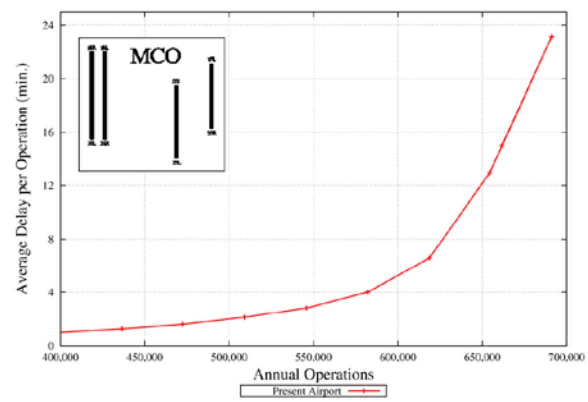

Figure 1. Annual Service Volume (ASV) Estimates

As illustrated in Figure 1, ASVs show that the expected average delays increase exponentially as the number of operations increases. Delays are based on the annualized number of operations in the schedule. Note that annualization of the data is necessary because the independent variable for an ASV is the annual number of operations. Once the ASV is calculated, its results are intuitive. One could interpret the resulting curve to indicate that the ultimate capacity of an airport is the point at which the ASV begins to increase very quickly with small increments in demand. This approach, however, has some limitations.

Limitations of Current Methods

Annualization of the data is one of the limitations. While it captures a variety of factors that relate to the operational performance of an airport, some of those factors vary seasonally, and delays are projected on a monthly basis. For some airports, delays are significantly higher in the summer and winter than in the spring and fall. For example, if airline schedules were to increase significantly in the spring, the annualization approach might significantly overestimate the delay impact. The annualization effect can be modest at some airports, but at a few airports, the size of the impact can be significant. For example, at San Francisco International Airport (SFO), delays are highly seasonal because persistent low-lying fog occurs only during certain times of the year. When fog is present, airport capacity is significantly reduced, leading to large delays. ASVs do capture this effect; however, it is factored in on an annual basis, leading to possibly substantial errors in monthly delay estimates.

A second limitation relates to the shapes of airline schedules by time-of-day. ASVs are calculated using a specific time-of-day profile for airport demand. While the latest information is used when the ASVs are computed, if the underlying schedule profile changes, without significantly changing overall demand, the ASV approach would not recognize any potential for change in delay. Airlines can adjust the shape of their schedules in response to a variety of changing economic circumstances in order to optimize their operations [6]. Figure 2 shows an example of arrival and departure schedules for two different days at ATL, where one is significantly more peaked than the other. Even though the more peaked schedule has fewer operations, it is more prone to delay than the less peaked schedule. An analysis based solely on ASVs would lead to an incorrect conclusion about delays for these two demand sets.

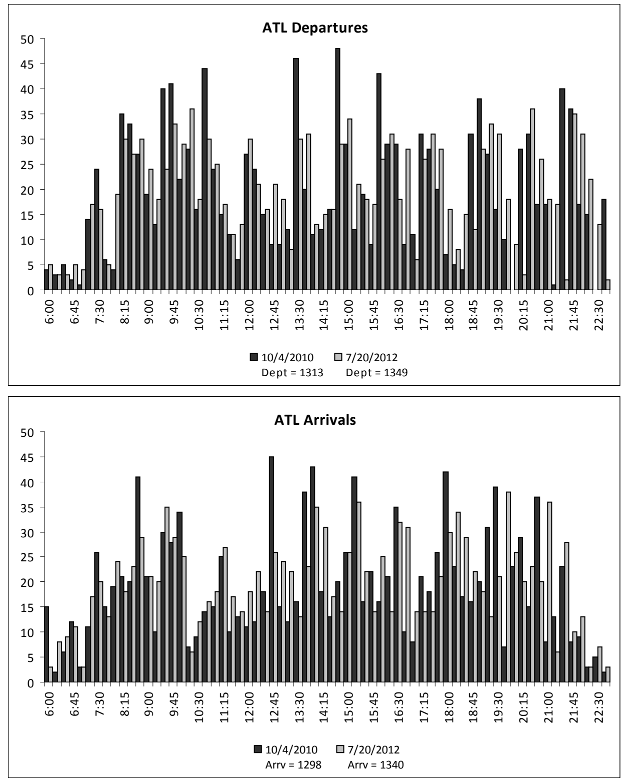

Figure 2. Scheduled Operations by Time of Day

A third limitation is due to delay propagation. ASV analysis focuses on queuing delays at individual airports. However, delays at one airport often propagate to other airports downstream. For example, Figure 3 shows that delays at St. Louis Lambert Field (STL) are better correlated with delays at John F. Kennedy International Airport (JFK) than they are with traffic at STL itself. While it is important to understand the first order delay dynamics at individual airports, it is also important to know their impact system-wide.

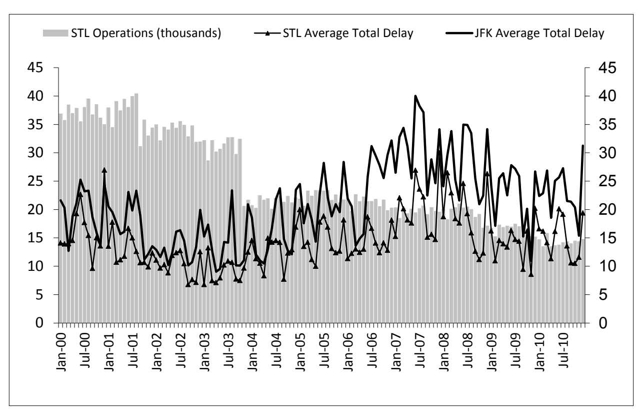

Figure 3. Operations and Delay History at STL

Addressing Modeling Limitations

The FAA addresses these three limitations with a detailed simulation using a NAS-wide model called SimCore. While SimCore is a very detailed and fully functional delay model, it has some limitations as well. It is not able to capture every aspect of the NAS that can affect delays, like convective weather, TMI's, GDP's, or flight cancellations. In addition, it is depends on detailed schedule data.

In addition to the above-mentioned limitations, each delay models depends on an accurate demand projection. Airlines release their schedules up to a year in advance of the day on which a flight is operated. While the detailed information for close-in schedules is very useful, there are issues with schedules more than a couple of months in advance of a flight. We have found that, the details of schedules more than three months out are less reliable than those of close-in schedules. Airlines often continue to optimize their schedules up to one to two months out due to changing economic conditions.

Taking into consideration the dynamic nature of airline schedules beyond three months out, we have collaborated with the FAA's Office of Aviation Policy and Plans (APO) to supplement the airline schedules with a demand forecast for each airport 4 to 12 months out. Since APO is responsible for generating the FAA's Terminal Area Forecast (TAF) [7], they have designed a methodology to provide the Performance Analysis office with a similar product focused on the near-term of 4 to 12 months, on a monthly basis. Although APO's near-term forecast provides us with a reasonable guide for overall near-term demand, it does not provide the level of detail that is included in a schedule. In particular, it does not provide flight-level details, but instead provides station level monthly estimates of demand. While the level of detail provided in the APO forecast data is not sufficient to run SimCore, it is sufficient for using an ASV or SDC type of model.

To take advantage of the strengths of each model and quality of available data, it seems most appropriate to use SimCore for projecting delays for months that are close-in, and using an ASV or SDC approach to model delays for months that are farther out in the future. Since the SDC approach addresses only the first limitation, it may be most suitably used for airports where seasonal variations matter. For airports where seasonal variations are not likely to be a factor in delays, the ASV method is completely adequate.

Seasonal Delay Curve

There are four fundamental steps to generating SDCs:

1. Identifying airports of interest
2. Grouping months into seasons based on similar Airport Arrival Rate (AAR - the published number of landings per hour that can be made at an airport based on specific conditions) and Airport Departure Rate (ADR - the published number of departures per hour that can be made from an airport based on specific conditions)
3. Estimating delay as a response to schedule levels by simulating the arrival and departure queues using the published AARs and ADRs for the identified seasons
4. Fitting a representative curve to estimated delays as a response to demand

Focus on Watch List Airports

We performed seasonal grouping of months for each of the Core 30 airports to test the validity of this approach and found that months at most airports could be grouped into two or three seasons. While applying this technique was useful in terms of validating the notion that airports have seasonal performance characteristics, we believe that it is practical to apply the seasonal delay analysis to airports where demand approaches capacity, e.g., the most delay prone airports. As such, we focus our attention on eight key airports that are among the busiest in the country and have demand that at times pushes up against the limits of their capacity. This is demonstrated by their historic delay profiles reported in the FAA's Operational Network (OPSNET) [8]. The airports that are of most interest are ATL, BOS, EWR, JFK, LGA, ORD, PHL and SFO. These airports were selected to be on the watch list because either they are the most prone to delays or because they handle such large volumes of demand. Figure 4 shows the monthly delay trends for these airports as reported in OPSNET.

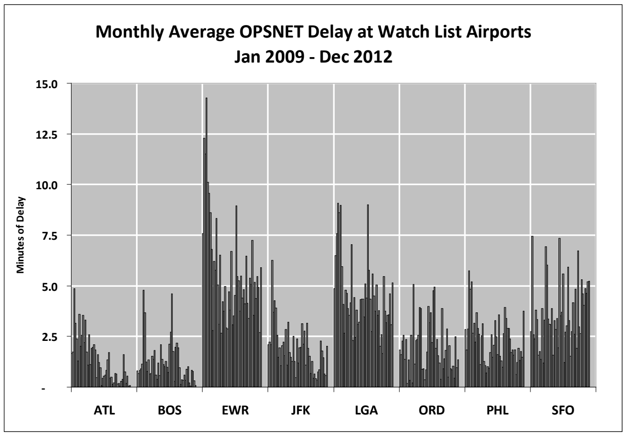

Figure 4. Monthly Average OPSNET Delays

Grouping Similar Months into Seasons

For each of the Core 30 airports, we identified months to group into seasons by applying pair-wise Kolmogorov-Smirnov tests on the sum of hourly called arrival rates (AAR) and departure rates (ADR). Initially the seasonal grouping analysis included all hours of the day. However, we found that it is more appropriate to apply the grouping analysis only for key operating hours at the airports, typically 6:00 to 23:00. Basing the grouping decision on the most relevant times of day made differing months more distinct from each other and helped identify similar months.

Kolmogorov–Smirnov Testing

The Kolmogorov–Smirnov (KS) test is a nonparametric test for comparing continuous, one-dimensional probability distributions. The KS test can be used to compare a sample to a reference probability distribution (one-sample KS test), or to compare two samples (two-sample KS test). The two-sample KS test is sensitive to differences in both location (mean and median) and shape (variance and skewness) of the empirical cumulative distribution functions (CDFs) of the two samples, making it a very compelling method for comparing two samples. [9]

To conduct the tests, we collected 5 years of hourly called rates (AAR+ADR) for the core 30 airports. Each hour was a data point within a monthly data set. We treated all five years of data for the same month as a single data set. The initial grouping of the five years of data generates 12 data sets, one for each month of the year. For each airport, we applied the two-sample KS test on all pair combinations of each airport's 12 monthly data sets.

The two graphs in Figure 5 and Figure 6 below show the CDFs of called rates for different months at SFO to illustrate how similar months and dissimilar months appear in the KS-test.

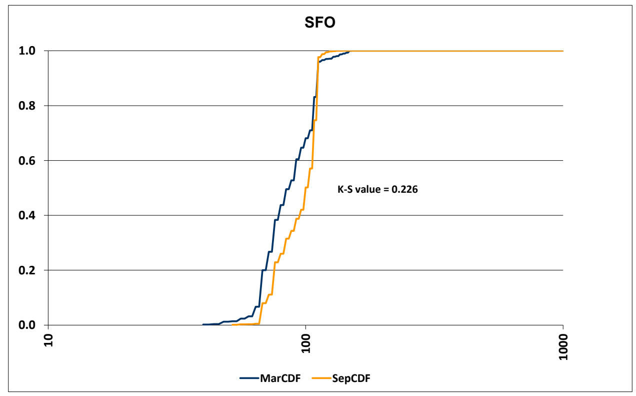

Figure 5. CDFs for Mar and Sep at SFO

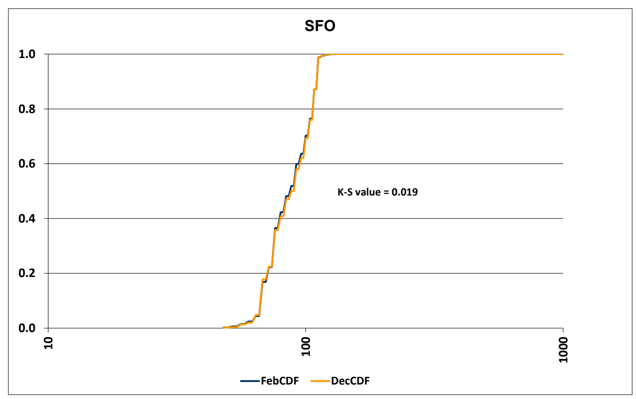

Figure 6. CDFs for Feb and Dec at SFO

We set thresholds for the KS testing at three levels: 0.075, 0.150, and 0.225. At some airports, 0.075 was practical to use, while at others, 0.150 made more sense. In the grouping exercise, some month pairs were allowed to exceed the selected threshold – primarily because variation is higher at some airports than others. However, we did not allow groupings to include pairs of months where the KS test exceeded 0.225. For example, if a successive string of months seemed to be similar, but one month in the sequence was a little bit off, we allowed it to remain within the season. Table 1 illustrates how the KS test results are used to group similar months for SFO.

Table 1. K-S Testing to Identify Similar Months

SFO	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Jan	0.00	0.05	0.07	0.13	0.12	0.14	0.22	0.18	0.22	0.22	0.18	0.05
Feb	0.05	0.00	0.05	0.10	0.11	0.13	0.21	0.17	0.22	0.21	0.17	0.02
Mar	0.07	0.05	0.00	0.12	0.12	0.10	0.19	0.15	0.23	0.19	0.18	0.05
Apr	0.13	0.10	0.12	0.00	0.05	0.05	0.16	0.11	0.19	0.16	0.14	0.09
May	0.12	0.11	0.12	0.05	0.00	0.05	0.16	0.11	0.19	0.16	0.14	0.09
Jun	0.14	0.13	0.10	0.05	0.05	0.00	0.12	0.08	0.15	0.12	0.11	0.12
Jul	0.22	0.21	0.19	0.16	0.16	0.12	0.00	0.05	0.04	0.00	0.10	0.20
Aug	0.18	0.17	0.15	0.11	0.11	0.08	0.05	0.00	0.08	0.05	0.07	0.17
Sep	0.22	0.22	0.23	0.19	0.19	0.15	0.04	0.08	0.00	0.04	0.09	0.20
Oct	0.22	0.21	0.19	0.16	0.16	0.12	0.00	0.05	0.04	0.00	0.10	0.20
Nov	0.18	0.17	0.18	0.14	0.14	0.11	0.10	0.07	0.09	0.10	0.00	0.16
Dec	0.05	0.02	0.05	0.09	0.09	0.12	0.20	0.17	0.20	0.20	0.16	0.00

In this particular example, SFO appears to have three distinct groups of months or seasons based on its AARs and ADRs: Apr – Jun, Jul – Nov and Dec – Mar. This seems intuitively correct, as SFO typically experiences morning fog in the winter and spring, but not as much in the summer.

Groupings for most of the Core 30 airports were reasonable, typically dividing the year into two to four seasons. In some airports looked like they would be better represented with more than four seasons, while others like SAN with just one season. To validate this part of the analysis we checked if airports in similar regions had the same seasonal groupings. Interestingly, this was not always the case. For example, the airports around New York City do not have identical seasonal groupings, but their groupings do have similarities. In addition, ORD and MDW appear to have very similar seasons, as do ATL and CLT. It appears that seasonal weather patterns may not be the only factor driving the seasonal groupings at the airports. There may be some local factors influencing seasonal capacity variations at airports, but it does seem that seasonal weather is the dominant factor. This is a useful validation step in the process. The two-sample KS test provided intuitively acceptable seasonal grouping results.

Table 2 shows a sample of the seasonal grouping results from KS testing AAR + ADR distributions for the watch list airports.

Table 2. Seasonality at Watch List Airports

	Spring	Summer	Winter
ATL		Jun - Oct	Nov – May
BOS	Mar - Jun	Jul - Nov	Dec – Feb
EWR		Jun - Oct	Nov – May
JFK	Mar - Jun	Jul - Feb	
LGA	Feb - May + Nov	Jun - Oct	Dec – Jan
ORD	Mar - May	Jun - Nov	Dec – Feb
PHL		Aug - Sep	Oct – Jul
SFO	Apr - Jun	Jul - Nov	Dec - Mar

Note that while LGA has three seasons, "Spring" actually represents a split season, Feb through May plus Nov. The rates used at LGA in Nov closely match the rates called in Feb through May. The seasons at other airports are composed of contiguous months.

Estimating Delays as a Response to Levels of Demand

Once airport capacity has been grouped by season, the next step is to estimate the delays at the watch list airports by season using a set of schedules that span a broad range of demand. This entails simulating arrival and departure queuing at the airports based on scheduled arrival and departure times.

The criteria for selecting these days was twofold. First, the schedules need to be relatively recent, to minimize influence of structural change within the industry. Secondly, the schedules need to span a broad enough range of demand to be conducive to fitting a response curve to the data. We collected a set of 21 daily schedules for each watch list airport. The days used to simulate demand for each airport are listed in Table 3.

Table 3. Scheduled Demand Used in Simulation

ATL		BOS		EWR		JFK	
Date	Total Ops	Date	Total Ops	Date	Total Ops	Date	Total Ops
20111124	1650	20111203	591	20101125	702	20101125	935
20110917	2117	20120101	706	20121027	872	20101106	1025
20120219	2218	20110226	781	20110312	950	20101204	1037
20120124	2430	20110423	842	20120129	1011	20110119	1056
20111004	2483	20110123	856	20110208	1124	20101005	1102
20110424	2487	20110313	912	20110209	1132	20101006	1109
20101212	2525	20111202	958	20110225	1137	20110109	1112
20101214	2585	20110424	964	20101004	1154	20101004	1114
20101017	2593	20111031	982	20101116	1163	20110323	1115
20110412	2596	20120313	1012	20101110	1182	20101018	1118
20101201	2604	20101215	1018	20101130	1182	20110101	1126
20101220	2621	20110704	1032	20101103	1186	20101130	1131
20101123	2657	20101104	1063	20110426	1211	20101001	1137
20101006	2662	20101007	1072	20110411	1214	20101122	1143
20101029	2670	20110328	1079	20110512	1217	20110418	1174
20101022	2687	20101014	1084	20110309	1219	20110102	1184
20101004	2693	20101202	1086	20101216	1252	20101216	1199
20101021	2718	20110407	1142	20110817	1253	20110618	1218
20110617	2764	20110520	1142	20110627	1254	20110726	1261
20110720	2779	20110713	1196	20110630	1288	20110812	1279
20120720	2789	20110812	1217	20120719	1311	20110623	1284

LGA		ORD		PHL		SFO	
Date	Total Ops	Date	Total Ops	Date	Total Ops	Date	Total Ops
20111225	525	20111124	1565	20101004	1289	20101125	767
20101106	610	20111008	1972	20101005	1268	20110319	927
20110430	634	20101120	2088	20101008	1326	20110116	999
20120728	731	20120131	2228	20101010	1214	20101214	1056
20110123	965	20101121	2304	20101111	1318	20101103	1086
20101107	984	20110111	2362	20101123	1336	20101222	1088
20120729	1016	20110918	2406	20101207	1252	20101012	1105
20120110	1125	20101123	2440	20101210	1244	20101021	1105
20111031	1128	20110113	2476	20101220	1294	20110516	1122
20101214	1156	20101109	2520	20110101	1131	20101011	1123
20101122	1165	20101210	2520	20110107	1233	20111023	1129
20101006	1175	20101118	2529	20110216	1283	20120501	1154
20101102	1179	20101005	2546	20110303	1333	20110519	1164
20101004	1181	20101006	2588	20110410	1199	20111005	1164
20101001	1182	20100612	2592	20110507	1058	20110627	1195
20101110	1184	20110824	2640	20110522	1160	20120503	1195
20101206	1192	20110509	2644	20110603	1313	20110710	1196
20110214	1194	20120730	2677	20110628	1351	20110706	1202
20101124	1197	20110519	2683	20111029	1003	20120913	1243
20110722	1202	20110629	2745	20111125	509	20120618	1278
20110210	1220	20110708	2769	20111227	1326	20120727	1303

We ran these schedules through a queuing simulation against five years of each season identified for each of the eight airports respectively. The queuing simulation called Airport Capacity and Slot Assignment Tool (ACASAT), developed by Metron Aviation for the office of Performance Analysis at the FAA. ACASAT is a stand-alone Java tool that models an airport as a single resource. The software is currently in Beta.

The simulations were run in batch mode for each season. We filtered the results to exclude extreme data points. Extreme points occur on days when the weather is particularly challenging. Since we are only concerned about excluding extreme points on the high side, we identified them as points that are more than 3 times the interquartile range away from the 75th percentile [10]. We took the average delay of all operations excluding the extreme values to represent the delay associated with a particular schedule for a given season. We did this for each of the schedules associated with an airport to generate a set of delay values each corresponding to a level of demand.

Since we simulated the delays with a queuing model, it is appropriate to use something similar to the functional form for a wait queue to fit the resulting data. Specifically, we used:

$W_q = \tau + \lambda/(\mu\ (\mu - \lambda))$, where τ represents a nominal minimum average delay, λ represents the arrival rate (e.g., number of scheduled operations per day), and μ represents the capacity of the airport (daily AAR + ADR). Note that if we subtract τ from W_q, the expression becomes the formula for expected waiting time spent in a queue for an M/M/1 system from classic queuing theory [11, 12]. We added τ to allow for a second parameter to help fit the function to the simulated results. One parameter is a location parameter (may be interpreted as the minimum possible delay time) and the other is a shape parameter (may be interpreted as the maximum capacity of the airport). ASVs are a two parameter function as well.

Note that one of the simplifying assumptions include that (after adjusting by τ) the airport behaves similar to an M/M/1 queuing system:

1. the airports are treated as single server queues
2. the scheduled operations can be represented as a Poisson process having rate λ
3. the service times are exponentially distributed with mean $1/\mu$

The advantage of making these assumptions is that it makes the analytical formulas easy to work with and interpret. A disadvantage is that in reality the Poisson / exponential assumptions may be violated. For example, the inter-arrival times may have dependencies, particularly at hub airports. This is an area that we could be explored further validate this method or to suggest an alternative method. For the purpose of this study, however, we will proceed with the above stated assumptions.

Fitting a Curve to the Data

The goal to curve fitting is to identify the parameters that make the curve fit the data well. In the functional form that we chose, our curve has two parameters to estimate. We used the root mean squared error (RMSE) to measure how well the curve fit the data and selected a value for the parameter that minimized this error.

Results

The results in Tables 4 through 6 show the value of the parameters that generate the best fit for the data for each airport along with a measure of the goodness of fit. The optimal parameter, μ, can be interpreted as an implied capacity of the airport. The optimal parameter, τ, can be interpreted as an implied minimum average delay at the airport. The minimized error term reflects a measure of how good the approximation can be in terms of average minutes of delay.

Table 4. Shape Parameter

μ: Implied Capacity (operations per day)			
	Spring	Summer	Winter
ATL		3059	3073
BOS	1323	1343	1291
EWR		1448	1431
JFK	1424	1397	
LGA	1263	1263	1258
ORD	3220	3276	3014
PHL		1696	1737
SFO	1507	1526	1375

Table 5. Location Parameter

τ : Implied Minimum Delay (min)			
	Spring	Summer	Winter
ATL		2.9	2.7
BOS	1.5	1.2	3.6
EWR		1.5	1.9
JFK	5.2	6.9	
LGA	5.0	5.8	6.5
ORD	1.5	1.5	2.1
PHL		4.6	4.4
SFO	1.5	1.5	2.8

Table 6. Goodness of Fit

RMSE (minutes of delay)			
	Spring	Summer	Winter
ATL		1.0	0.9
BOS	0.9	0.8	2.0
EWR		0.6	0.8
JFK	1.4	1.8	
LGA	3.6	4.8	4.2
ORD	0.3	0.3	0.6
PHL		0.5	0.5
SFO	0.3	0.4	1.5

Note that some airports have a higher error value, which generally corresponds with an overall higher average amount of delay at that airport. This effect on error is demonstrated by the three SDCs derived for SFO, as shown in Figures 7 through 9.

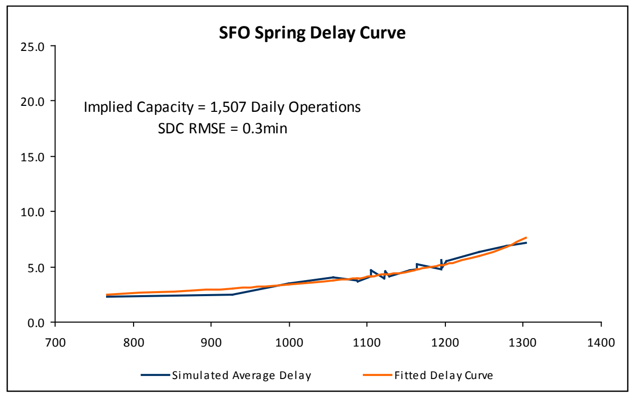

Figure 7. Estimated Delays at SFO for Spring

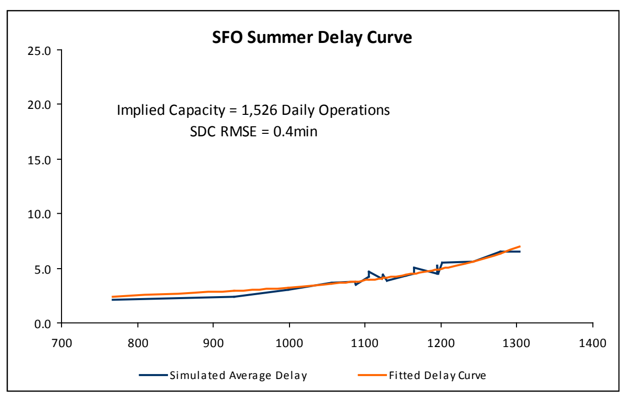

Figure 8. Estimated Delays at SFO for Summer

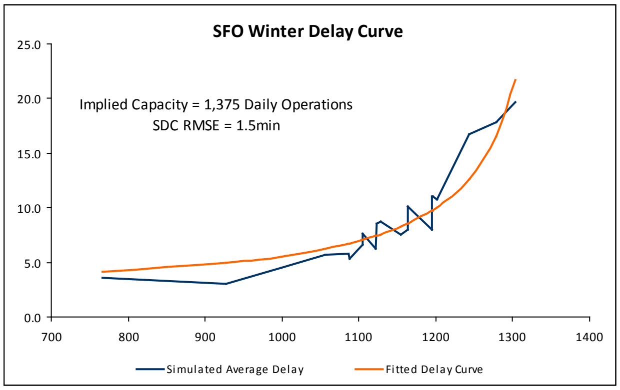

Figure 9. Estimated Delays at SFO for Winter

Schedule Anomalies

The simulation results for summer at ATL shown in Figure 10 display an unexpected reduction in delay at the very high end of the demand range. This type of result can occur when the shape of the schedule changes. As an example, the charts in Figure 2 show the frequency of arrivals and departures at ATL in 15-minute time bins for two of the schedules used to simulate demand at ATL. Even though the schedule shown in the dark columns (10/4/2010) has about 100 fewer operations than the schedule in the lightly shaded columns (7/20/2012), the schedule for 10/4/2010 is more peaked and results in about 2 minutes more delay per flight on average. This comparison demonstrates how carriers can sometimes help mitigate delays at their hubs without reducing their level of operations, and in some cases, can even increase the number of operations with no adverse effect to delays on average.

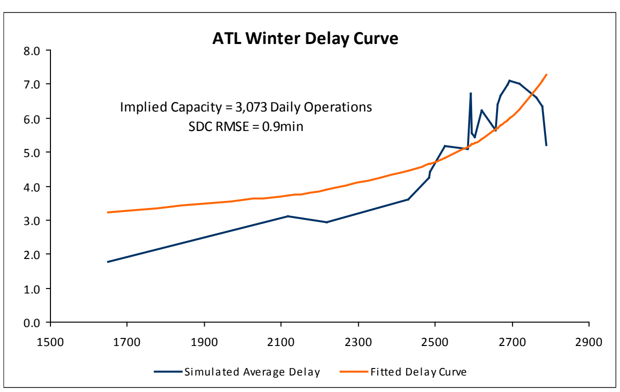

Figure 10. Estimated Delays at ATL for Winter

This result adds to the challenge of forecasting delays. Currently we are including the schedule data in our sample without regard for its shape with the assumption that we have a sufficient number of data points to represent a reasonable range of possible schedule shapes. Thus, the resulting delay estimates are reasonable especially when considering that we are applying this method to schedules that are farther out in the future and their final shape is still uncertain.

Validation

The SDC model fits the simulated data results quite well as shown in Table 7. In fact, the delay curves generated by the SDC method fit the simulated data better than the ASV. Table 7 shows the improvement in the RMSE between using SDC and ASV for predicting delays at the watch list airports, indicating an improvement for each season at each airport.

Table 7. Comparison of Goodness of Fit

	Spring	Summer	Winter
ATL		-0.4	-0.3
BOS	-2.2	-1.5	-4.8
EWR		-8.3	-7.2
JFK	-3.4	-0.9	
LGA	-2.5	-2.7	-4.1
ORD	-3.6	-3.7	-1.5
PHL		-1.2	-1.4
SFO	-8.2	-8.4	-3.9

While these results are encouraging, we ideally want to validate the results against true operational data. Three sources of operational data for validation are OPSNET, ASPM and U.S. Department of Transportation (DOT)'s Airline On-Time Statistics.

By design, OPSNET tends to understate delays because of its 15-minute rule and reporting only delays that result from Air Traffic Control (ATC). If an operation is delayed less than 15 minutes or for other than ATC reasons, it is not considered delayed within OPSNET. In fact, airports can have have a number of days with zero reported delay within OPSNET. On the one hand, the 15-minute limit may hinder the comparison, but the ATC reason for delay is probably consistent with the type of delays that we are trying to model.

Alternatively, ASPM data provides two different types of delay estimates. Its data allows us to compare a flight's actual performance to an idealized duration of flight, or to its scheduled arrival time. An idealized duration of flight can vary significantly from what is in the published schedules which may include some padding. ASPM delay estimates based on nominal flight times can over estimate delays when compared to delays vs. schedules.

The third source for validation is the DOT's Airline On-Time Statistics [14]. One draw back with this data set is that not all carriers are included. Airline on-time data cover the 14 U.S. air carriers that have at least 1 percent of total domestic scheduled-service passenger revenues, plus one other carrier that reports voluntarily.

Table 8 provides validation results comparing SDC delay estimates with each of the above mentioned data sources for each of the watch list airports. The values listed are the correlation coefficients from comparing SDC estimated delays to each of the corresponding data sets.

Table 8. SDC Correlations

	% Delayed	Schedule Delay	Nominal Delay	OPSNET Delay
ATL	0.44	0.50	0.54	0.43
BOS	0.25	0.32	0.39	0.01
EWR	0.61	0.54	0.59	0.58
JFK	0.59	0.59	0.51	0.60
LGA	0.30	0.34	0.36	0.18
ORD	0.58	0.60	0.56	0.53
PHL	0.32	0.13	0.18	0.14
SFO	0.35	0.49	0.53	0.51

The correlations shown in Table 8 are calculated using data from 2008 to 2012. In general, the ASPM data sources provide the best validation results, with modeled delay estimates for ATL, EWR, JFK, ORD, and SFO having the strongest correlations.

Three factors that may weaken the correlations shown above include severe weather events, changes to the shape of the schedule over time, and inaccuracies in the data used in building the SDC model. Schedule shapes can change over time if a dominant carrier changes its scheduling paradigm or if there is a significant share shift among carriers at the airport.

Future Work

We are currently working to develop a better understanding of how to use long-term weather forecasts and significant weather events to enhance our delay prediction capabilities. Significant weather event forecasts could enhance the delay forecast capability for each of the models currently in use today.

We are also exploring different functional forms to represent the delay curves. While it is desirable to use functional forms with interpretable parameters, one feature the current functional form lacks is going through (0, 0). That is a desirable feature because it would be consistent with the intuitive expectation that as operations go to zero, average delays go to zero as well.

We have also seen that schedule shape can be a significant factor in determining delays. Analysis of how to measure the shape or the degree to which a schedule is peaked may be helpful in forecasting delays as well.

Conclusion

Delay models are not perfect, but we are continually working to refine the models we use. The SDC method, based on seasonal called rates at airports, seems to be a good complement to the other delay models currently in use, providing the analyst with a richer toolset to estimate delays. We plan to further refine the SDC models and use them to enhance the delay forecasts reported in the monthly delay report.

References

[1] Federal Aviation Administration (2009). FAA Flight Plan 2009–2013. [Online] Available from: http://www.faa.gov/about/plans_reports/media/flight_plan_2009-2013.pdf [Accessed February 2013].

[2] Section 413 of the FAA Reform and Modernization Act of 2012, 112th Congress Public Law 95, Government Printing Office. [Online]

Available from http://www.gpo.gov/fdsys/pkg/PLAW-112publ95/html/PLAW-112publ95.htm [Accessed March 2013]

[3] Federal Aviation Administration (March 2013). Monthly Delay Report. Internal report prepared by Federal Aviation Administration, Office of Performance Analysis, Washington, DC.

[4] Federal Aviation Administration (2012). Annual Service Volume Study. Internal report prepared by Federal Aviation Administration, FAA William J. Hughes Technical Center, Atlantic City International Airport, New Jersey.

[5] Chin, Meilus, Murphy, and Thyagarajan, (2013) "Modeling and Managing Airport Performance – Chapter 4: Forecasting Airport Delays", Federal Aviation Administration, Office of Performance Analysis, Washington, DC, Wiley.

[6] Petroccione, L. (May 7, 2007). "Delta's Operation Clockwork Transforming the fundamentals of an airline", Air Transport World Online, [Online] Available from: http://atwonline.com/whitepapers/deltas-operation-clockworktransforming-fundamentals-airline-0215 [Accessed February 2013].

[7] Federal Aviation Administration (2013), Terminal Area Forecast, [Online] Available from http://aspm.faa.gov/main/taf.asp, [Accessed March 2013]

[8] Federal Aviation Administration, The Operations Network (OPSNET), FAA Operations & Performance Data [Online] Available from https://aspm.faa.gov/opsnet/sys/main.asp, [Accessed March 2013]

[9] Mann, Schafer and Singpurwalla (1974), "Methods for Statistical Analysis of Reliability of Life Data", Wiley

[10] NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, [Online] Available from http://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm [Accessed March 2013]

[10] Gross and Harris (1985), "Fundamentals of Queuing Theory", Second Edition, Wiley

[12] Ross, Sheldon M., (1989). "Introduction to Probability models", Fourth Edition, Academic Press inc.

[13] Operational Data Reporting Requirements (Oct 2009). Federal Aviation Administration, Air Traffic Organization, Washington, DC.

[14] Research and Innovative Technology Administration, US Department of Transportation, Washington, DC. [Online] Available from http://apps.bts.gov/xml/ontimesummarystatistics/src/index.xml [Accessed March 2013]

2013 Integrated Communications Navigation and Surveillance (ICNS) Conference
April 23-25, 2013