

WEATHER NORMALIZATION FOR EVALUATING NATIONAL AIRSPACE SYSTEM (NAS) PERFORMANCE

Mark Hansen, Jing Xiong, University of California at Berkeley, NEXTOR,
Berkeley, CA, USA

Abstract

The need to benchmark air traffic management performance, predict future performance, and improve our understanding of air traffic operations has created a growing literature on the statistical modeling of delays in the National Airspace System (NAS). This paper contributes to that literature in three distinct ways. First we introduce an innovative delay metric that avoids the distortions created by schedule padding and can be decomposed into different flight phase components. Second, we examine closely how daily variation in weather-impacted traffic affects operational performance. Third, we consider the impacts of weather forecast errors as well as realized weather. Our results demonstrate the value of the new delay metric, show that simple daily averaging adequately captures the effect of weather-impacted traffic, and reveal that false-positive weather forecast errors are a significant source of delay in the NAS.

1. Introduction

Delay is a common occurrence in the National Airspace System. Imbalance between airport demand and capacity is a primary cause of delay. Convective weather is also a major contributor. Air traffic management initiatives, whose goal is to manage the flow and mitigate the safety and efficiency impacts of delay-capacity imbalances, sometimes make the delay worse. We cannot reliably evaluate the effectiveness of air traffic management actions without normalizing for weather, traffic, and other factors. Thus we need understand the contribution of these factors both to assess the performance of current air traffic flow management and develop strategies for improving it. Econometric modeling of NAS operational performance is a promising means of developing such understanding.

There is a large growing literature on this subject. More recent contributions include Hansen et al. [1], who built an econometric model to capture the impacts of terminal weather, convective weather, arrival queuing, seasonal effects, and secular effects on delay. Hsiao et al. [2] extends the

model and investigate the time of day effects of arrival queuing, the effects of scheduled arrivals and most importantly, the interaction between scheduled arrivals and weather conditions. Post et al. [3] has a similar regression model built to analyze delay. The dependent variable in his paper is total delay, calculated as the sum of departure, taxi-out, en route, and taxi-in delay. The exogenous variables are (1) the number of scheduled arrivals during instrumental approach condition; (2) the excess of scheduled arrivals over the recorded airport acceptance rate; (3) an en-route weather index; (4) a dummy variable for weekend effects.

Theoretically, convective weather itself will not cause the delay if the traffic is very low. Callahan et al.[4] introduced the concept of WITI (Weather Impacted Traffic Index) to statistical delay modeling. After dividing national airspace into an I by J grid, traffic counts in region (i,j) at time t , $T_{i,j}(t)$, and severe convective weather incidence in region (i,j) at time t , $W_{i,j}(t)$, are obtained. Using these variables one can compute how many aircraft are exposed to bad weather at time t in the system: $\sum_{i,j} W_{i,j}(t) \cdot T_{i,j}(t)$. These results

can be aggregated over time to create an hourly or daily WITI. Sridhar [5][6] furthers the research on WITI by characterizing its time series over the course of a day and using traffic counts on a weather-free "reference" day to calculate $T_{i,j}(t)$. By basing the WITI on a reference day, Sridhar measures the coincidence of weather and traffic in the absence weather-related disruptions. As an additional refinement, Sridhar expands the regions around severe weather that are designated weather-impacted to 20 miles.

The purpose of this paper is to improve upon existing WITI-based delay regression models in three specific ways. First, we employ two delay metrics: average delay against schedule and a daily average flight time index (DAFT). The former measures the performance in terms of adherence to schedule. The latter avoids the distortions resulting from schedule padding while also allowing an improved understanding of how causal factors affect delay in different phases of flight. Second, we propose and implement an innovative moment-

based characterization of the daily WITI time series. Third, we consider the effects of forecast weather as well as realized weather in our model. This contribution is particularly critical since most air traffic management actions are based on anticipated weather.

The remainder of this paper is organized as follows. Section 2 describes the delay metrics and explanatory variables employed in this study. Section 3 specifies four delay models and presents estimation results. Section 4 highlights our main findings and makes suggestions for future research.

2. Model Description

We first propose the general econometric model we would use to estimate the delay.

$$D(t)=f(W(t), V(t), I(t), B(t)) + \nu_t$$

Where:

$D(t)$ is some NAS delay metric for day t ;

$f(\cdot)$ is a deterministic function;

$W(t)$ is a vector characterizing the WITI in day t ;

$V(t)$ is average wind speed at major airports in day t ;

$I(t)$ is proportion of flights schedule to land under IFR conditions in day t ;

$B(t)$ is a vector capturing the weather forecast results in day t ;

ν_t is stochastic error term;

Description of variables in each category follows.

2.1 Delay Metrics

2.1.1 ASPM 75 Daily Average Delay

As part of its effort to monitor and analyze the operation performance of the aviation system, FAA maintains a web-accessible data base known as Aviation System Performance Metrics (ASPM) [7]. Our Daily Average Delay metric is based on performance data in 75 domestic airports and obtained from ASPM. The arrivals counted in ASPM exclude General Aviation, Military, and a small number of commercial flights. We used ASPM to calculate daily average delay—the total arrival delay in a specific day divided by the total arrivals completed in that day. This metric treats negative delay values as 0's.

2.1.2 Daily Average Flight Time Index (DAFT)

ASPM delay compares the actual arrival time with the scheduled arrival time. Scheduled flight times are adjusted to reflect anticipated delay. Such buffering may distort ASPM delay metric. Buffering also masks variations in the ASPM delay metric because ASPM only counts delays that exceed the schedule buffer. We developed an alternative metric, not affected by such buffering or the exclusion of negative delays, based on day-to-day variation in average flight times. We refer to this as the Daily Average Flight Time Index (DAFT).

The DAFT is computed from the ASQP data base, which contains scheduled arrival and departure times, as well as actual Out-Of-On-In times, for every domestic flight operated by large US carriers. The methodology of developing DAFT metrics is as follows:

- Identify a set of OD pairs for which there are flights in ASQP for 95% of the days in the analysis period (2001-2006 in this case).
- For each of these OD pairs, compute the daily average flight time between scheduled departure and actual arrival time as well as averages for its four components: departure delay, taxi-out time, airborne time, and taxi-in time.
- For each of these OD pairs, average the daily average over all the days in the analysis period.
- For each OD pair and day, find the deviation between the daily average and the average over all days.
- For each day, compute the daily average deviation over all OD pairs, weighted according to the total number of flights for each OD pair over the analysis period.
- Adjust the daily average to account for the small number of OD pairs that may be missing on any given day.

The Daily Average Flight Time Index (DAFT) metric contain five variables:

$DDDL(t)$: in day t , the average deviation of departure delay;

$DTOUT(t)$: in day t , the average deviation of taxi-out time;

$DAIR(t)$: in day t , the average deviation of airborne time;

$DTIN(t)$: in day t , the average deviation of taxi-in time;

$DTFT(t)$: the sum of the previous four components, and is the average deviation of total flight time from scheduled departure to actual arrival for day t .

The DAFT metric does not depend on scheduled block time and is therefore unaffected by schedule padding. Moreover the four-component DAFT metric allows more detailed investigation about how different causal factors affect operational performance. For example, does convective weather cause delay by increasing airborne time or instigating more ground holding?

2.2 Traffic Demand

We use the scheduled arrival counts (in one hundred thousands) to represent the traffic demand in a specific day. We expect higher traffic to increase average delay, all else equal. We use the number of scheduled arrivals for the same 75 airports that form the basis for the average delay metric, based on ASPM data.

2.3 WITI

Since the concept of WITI has been introduced by Callaham et al.[4], several studies have sought to improve our understanding of the relationship between WITI and delay. The WITI metric depends on how many aircrafts are impacted by severe weather conditions at a given time. While this is clearly an important explanatory variable, how to incorporate it into a statistical model remains an open issue.

Sridhar et al. [5] [6] have developed the WITI metric by multiplying every-minute traffic data for a reference day with good weather (from ETMS) with observed every-five-minute convective weather data (from NOWRAD). After the one minute WITI metric is generated, statistical features, such as WITI histogram and time domain, are proposed to capture the characteristics of WITI distribution through the day.

We extend the research work reported by Sridhar and use the same one minute WITI data used in his models to develop a new statistical way to characterize the daily WITI time series. Each day includes 1440 one-minute WITI values. Our data set consists of 189 days with this WITI information. Figure 1 shows WITI time series for two selected days.

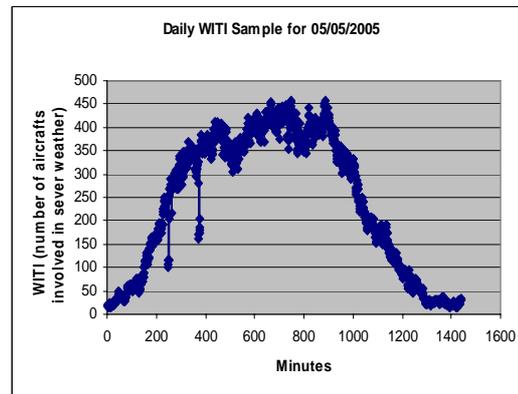
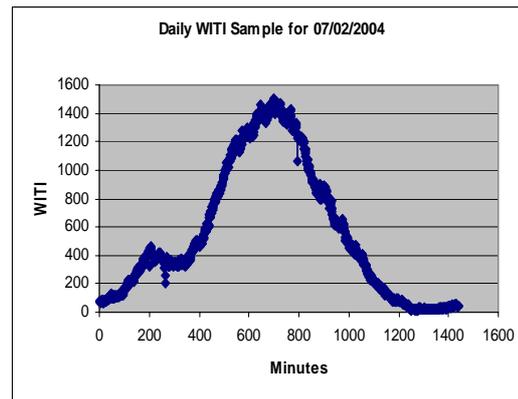


Figure 1. Example WITI Time Series

To characterize a daily time series of WITI values we define two random variables. The first one is the WITI value itself. We characterize the WITI value distribution on a given day in terms of its first three moments, specifically the mean ($W1$), standard error (standard deviation divided by the mean— $W2$), and skewness (a dimensionless measure of distribution asymmetry based on the third moment— $W3$). These moments characterize the magnitude of WITI values but not their time variation. We thus define another random variable which is the time of day when a particular encounter between a flight and a storm occurs. We also characterize the distribution of this random variable with three moments: the mean ($T1$), standard error ($T2$), and skewness ($T3$). This $T1$ measures the central tendency of how late in the day convective weather impacts occur, while $T2$ assesses the dispersion of weather impacts over the day. In summary, we generate six variables to characterize the distribution and time variation of WITI for a given day. This permits systematic study of which WITI features contribute to delay.

We notice that WITI values used by Sridhar et al. [5] [6] are based on weather-free reference day traffic only. In order to capture the interaction between WITI and the real traffic, we scale WITI mean ($W1$) based on reference-day traffic by

multiplying it by scheduled arrivals. Only WI is scaled in this way because all other WITI features are independent of overall traffic levels.

The WITI metric is assembled during the convective season, which includes July, September and October in 2004, as well as April, May, June, July and August in 2005. The final time series of WITI consists of 189 days observation. Therefore, all other variable metrics are developed for those 189 days.

2.4 Terminal weather

When airports are operated under instrument conditions or windy conditions, airport capacities are usually reduced. Thus we include wind speed and the incidence of IFR conditions in our delay models to represent the terminal weather condition.

2.4.1 Wind

For a specific day, we define variable Wind as:

$$Wind = \frac{\sum_{i,j} Windspeed_{i,j} * Arr_{i,j}}{\sum_{i,j} Arr_{i,j}}$$

Where,

i : time index, 96 quarters in a day

j : airport index, 75 airports in the country

$Windspeed_{i,j}$: wind speed (in knots) in quarter i , at airport j

$Arr_{i,j}$: number of scheduled arrivals minus cancelled flights, in quarter i , at airport j

From ASPM 75 quarterly operation data, we aggregate the wind speed to a daily average, weighted by the number of arrivals. This represents the average wind that would be faced by a flight if it arrived at its destination airport at its scheduled time.

2.4.2 IFR

For a specific day, we define variable IFR as:

$$IFR = \frac{\sum_{i,j} MC_{i,j} * Arr_{i,j}}{\sum_{i,j} Arr_{i,j}}$$

Where,

i : time index, 96 quarters in one day

j : airport index, 75 airports in the country

$MC_{i,j}$: in quarter i , at airport j , if airport is under IFR condition, $MC_{i,j}=1$, Otherwise $MC_{i,j}=0$

$Arr_{i,j}$: number of arrivals minus cancelled flights, in quarter i , at airport j

Based on ASPM 75 quarterly IFR information and weighted by arrivals, we aggregate the IFR to daily average. IFR is the fraction of flights destined for the ASPM 75 airports that would encounter IFR conditions if they landed at their scheduled times.

2.5 Weather Forecast Performance Metrics

The air traffic management uses weather forecast information to help decide if certain initiatives should be carried out. For instance, if convective weather is predicted to happen in two hours in Chicago, the flights designated to Chicago might be held on the ground at origin airports. Weather forecasts are essential for air traffic management to effectively mitigate capacity-demand imbalances. However, weather forecasting is not perfect, under-forecasting or over-forecasting can cause suboptimal decisions that increase delay. We anticipate that over-forecasting will lead to unnecessary TFM actions that increase delay. In the under-forecasting case, it is not certain if this effect is positive or negative on delays. It may force long, unanticipated reroutes, but also will reduce ground delays, even though in this instance such delays would have been advisable. The Collaborative Convective Forecast Product (CCFP) maintained by the Aviation Weather Center [8] has verification statistics for CCFP forecasts. CCFP forecasts each day during the convective weather season. Forecasts are issued every two hours, and each forecast identifies regions of the continental US where convective weather may occur 2, 4, and 6 hours after the forecast issuance time. The statistics are based on the counts of 40x40 km squares covering continental US. The forecast tabulates the number of cases for each of the four cases identified below:

Forecast	Observation	
	No	Yes
No	NN	NY
Yes	YN	YY

As NN event occurs whenever a 40x40 km square is not forecast to have convective weather and does not, while a YY event is counted when convective weather is realized in a square for which it is forecast. When convective weather occurs in square where it is not forecast, a NY event is counted, while a YN event involves a square in which convective weather is forecast but does not materialize. In our subsequent discussion, we will refer to NY events as “false-negatives” or

“under-forecasting” and YN events as “false-positives” or “over-forecasting.” For this research we consider aggregate counts of events over all forecast issuance times and time horizons. Thus weather in a given square at a given time is compared with three different forecasts 2, 4, and 6 hours prior.

An alternative index to describe the forecast result is Bias, defined as the ratio of the number of Yes forecasts to the number of Yes observations.

$$Bias = (YY + YN) / (YY + NY).$$

Bias measures the tendency of the CCFP in a given day toward over-forecasting or under-forecasting. When Bias is greater than 1, it is over-forecasting. While Bias is measured on a continuous scale, it is likely the variation in bias values over 1 (indicating the degree of over-forecasting) will affect delay differently than variation in bias values less than 1 (indicating the degree of under-forecasting). Thus we generate two variables Over-forecasting Ratio and Under-forecasting-Ratio based on Bias index.

$$\text{Over-forecasting Ratio} = \begin{cases} Bias - 1, & \text{if } Bias > 1, \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Under-forecasting Ratio} = \begin{cases} 1 - Bias, & \text{if } Bias \leq 1, \\ 0 & \text{otherwise} \end{cases}$$

The Over-forecasting Ratio coefficient reflects how over-forecasting affects delay and we expect it to be positive. The coefficient on Under-forecasting Ratio captures the delay impact of under-forecasting, which may increase or decrease delay, as discussed below. We estimate models with the two ratios and, alternatively, the absolute numbers of NY and YN events (in hundreds of thousands) respectively.

3. Model Specification and Estimation Result

3.1 Model Specification

The delay model is formulated as:

$$D(t) = \alpha + \lambda S(t) + \sum_{i=1}^3 \beta_i W_i(t) + \sum_{i=1}^3 \chi_i T_i(t) + \delta V(t) + \varepsilon I(t) + \sum_{i=1}^2 \phi_i M_i(t) + \sum_{i=1}^2 \varphi_i N_i(t)$$

$\alpha, \lambda, \beta_i, \chi_i, \delta, \varepsilon, \phi_i, \varphi_i$ are coefficients be estimated; t indexes the observation day; D refers to some delay metric--either ASPM average delay or one of the DAFT metrics; and all explanatory variables are as summarized in table 1.

Symbol	Variables	Definition	Mean	Std Dev	Sum	Minimum	Maximum
ASPM	ASPM Delay	Average daily delay in 75 airports, in minutes	13.196	5.688	2494.000	4.956	29.042
DTD	Total Delay	Average daily deviation of total flight time comparing to average, in minutes	0.903	6.986	170.671	-11.049	22.661
DDDL	Departure Delay	Average daily deviation of departure delay comparing to average, in minutes	1.265	5.688	239.097	-7.894	20.138
DTOUT	Taxi-out Delay	Average daily deviation of taxi-out time comparing to average, in minutes	0.035	1.102	6.537	-2.387	3.490
DAIR	En-route Delay	Average daily deviation of enroute time comparing to average, in minutes	-0.553	0.983	-104.466	-2.471	1.931
DTIN	Taxi-in Delay	Average daily deviation of taxi-out time comparing to average, in minutes	0.156	0.329	29.503	-0.440	1.899
S	Scheduled Arrivals	Daily Scheduled arrival counts, in 100,000	0.235	0.016	44.441	0.181	0.254
W1	WITI Mean		69.907	35.949	13212.000	4.681	176.800
W2	WITI Std. Err.		0.774	0.103	146.372	0.537	1.007
W3	WITI Skewness		0.393	0.302	74.276	-0.289	0.943
T1	Time Mean		678.062	42.921	128154.000	512.633	818.250
T2	Time Std. Err.		0.387	0.046	73.227	0.305	0.543
T3	Time Skewness		-0.114	0.155	-21.515	-0.542	0.417
V	Wind	Average daily airport wind speed weighted by number of arrivals, in knots	7.846	1.091	1483.000	5.186	11.393
I	IFR	Average daily proportion of time that airport operated under IFR	0.119	0.075	22.478	0.009	0.480
M1	False-positive	Counts of YN(When weather forecast as Yes, observed as No), in 100,000	0.401	0.282	75.716	0.000	1.079
M2	False-negative	Counts of NY(When weather forecast as No, observed as Yes), in 100,000	0.390	0.239	73.650	0.002	0.905
N1	Over-forecasting Ratio	Refer to the metrics discription in section 2	0.058	0.086	10.941	0.000	0.369
N2	Under-forecasting Ratio	Refer to the metrics discription in section 2	-0.135	0.212	-25.525	-1.000	0.000

Table 1. Summary of All Variables and Statistics

3.2 Estimation Procedure

We use Ordinary Least Square (OLS) to estimate numerous models based on different combinations of explanatory variables. However we will only present four models in this paper. The

first three models feature ASPM average delay as the delay metric. Model 1 includes all explanatory variables in category traffic demand, WITI and terminal weather. We only include two variables in weather forecast results category, the Over-forecasting Ratio and Under-forecasting Ratio. Model 2 is identical to Model 1, except for the weather forecast variables, which are the False-

positive and False-negative counts rather than the Over/Under-forecasting Ratios. From these two models, we select the one that yields the most satisfactory results. For the selected model, we observe some insignificant variables and eliminate them to formulate the more parsimonious model 3. We estimate model 3 and finalize the structure. Finally, we apply the specification in model 3 to DAFT delay metric (Model 4) including the total time metric and its four components.

3.3 Model Estimation Results

Table 2 summarizes the estimation results for all three ASPM models, the DAFT model, which includes the Total sub-model and the four component sub-models. The first three ASPM models and DAFT Total sub-model have R^2 's of .62 to .66. It should be noted that these results strictly comparable of the many other models in the literature which model total rather than average delay. Such models will tend to have higher R^2 's because total delay will obviously increase with the number of flights, and aggregate models get credit for this effect in their R^2 . The DAFT component sub-models have somewhat less explanatory power, with R^2 's in the 0.5-0.6 range.

Figures 2 and 3 show plot actual vs predicted values for ASPM model 3 and the DAFT Total sub-model. There is some curvature in the plots reflecting the limitations of a first-order model that excludes interaction and higher order polynomial terms. Nonetheless, degree of scatter is fairly consistent over the whole range of predicted values, suggesting that assumption of identically distributed errors is reasonably correct, thus

validating statistical inferences based on the OLS results.

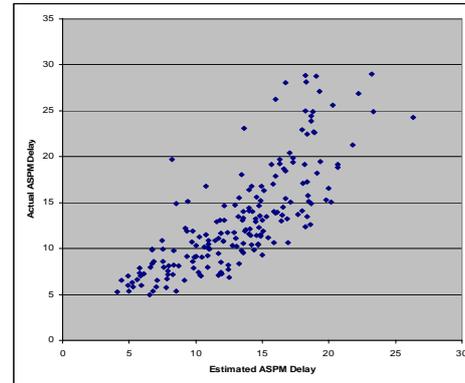


Figure 2. Scatter Plot of ASPM Delay Model 3

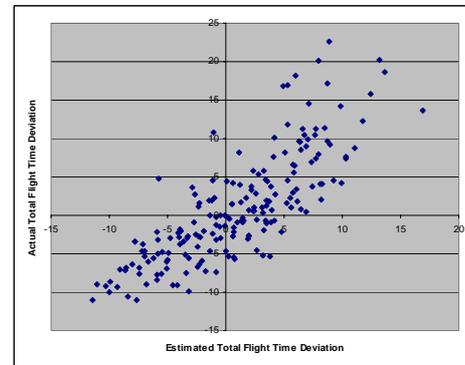


Figure 3. Scatter Plot of Total Delay Sub-model in DAFT Delay Model 4.

Category	Variable	ASPM Models						DAFT Model 4									
		ASPM delay 1		ASPM delay 2		ASPM delay3		Total		Departure		Taxi-out		Enroute		Taxi-in	
		Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.								
	Intercept	-76.079	32.061	-73.164	31.733	-16.009	4.519	-45.847	5.268	-27.402	4.555	-6.121	0.879	-12.048	0.817	-0.277	0.301
WITI	Scheduled Arr.	40.918	20.108	48.924	21.775	56.879	17.308	127.067	20.176	72.602	17.447	18.961	3.367	34.707	3.129	0.787	1.153
	WITI mean	0.104	0.011	0.084	0.019	0.065	0.013	0.065	0.015	0.057	0.013	0.011	0.003	-0.007	0.002	0.004	0.001
	WITI Std. Err.	29.283	12.571	27.921	12.377												
	WITI Skewness	-5.310	2.588	-5.367	2.559												
	Time mean	0.036	0.021	0.032	0.021												
	Time Std. Err.	52.712	28.112	50.350	27.640												
	Time Skewness	-0.995	2.763	-0.825	2.715												
Terminal	Wind	0.529	0.277	0.639	0.272	0.834	0.252	0.798	0.294	0.389	0.254	0.016	0.049	0.415	0.046	-0.022	0.017
	IFR	15.109	4.234	16.431	4.236	18.279	3.878	21.062	4.520	13.709	3.909	2.726	0.754	4.458	0.701	0.169	0.258
	False-positive			10.068	2.870	6.387	1.544	8.870	1.800	7.302	1.557	1.110	0.300	0.179	0.279	0.279	0.103
	False-negative			-6.841	4.542												
	Over-forecasting Ratio	12.349	3.609														
	Under-forecasting Ratio	-1.121	1.557														
	R-squared	0.624		0.636		0.617		0.655		0.611		0.614		0.581		0.491	

Values in **bold italics** are significant at 5% level
 Values in **bold** are significant at 10% level

Table 2. Model Estimation Results Summary

First, Scheduled Arrivals is positive and significant in all models except DAFT Taxi-in sub-model. This implies delay is sensitive to overall traffic in the NAS, even on a day free of convective weather. Interestingly, the coefficient on Scheduled Arrivals is much higher in the DAFT Total sub-model, suggesting that the much Schedule Arrival effect is masked in the ASPM models by schedule

padding and the truncation of negative delays to zero in ASPM delay. Based on the range of Scheduled Arrival values shown in Table 1, the difference between delay on a high traffic and a low traffic day is about 4 minutes based on the ASPM

model 3 and 9 minutes based on the DAFT Total sub-model.¹

Second, the WITI mean (*WI*) coefficients in three ASPM delay models and DAFT Total delay sub-model are statistically significant and positive. This implies, not surprisingly, that a higher mean WITI results in higher observed delay in the system. The magnitude of the effect is the same whether ASPM average delay or the DAFT Total metric is the dependent variable. Note that, as constructed here, the WITI value for a given convective weather pattern is proportional to Scheduled Arrivals. Thus increases in scheduled flights affect delay both directly and by increasing the WITI mean.

The DAFT component models provide a more nuanced picture of how WITI affects delay. It increases time spent at the origin airport (departure delay and taxi-out time) and also slightly increases taxi-in time, but has a negative effect—small but statistically significant—on en route times. The origin airport effects derive from a combination of TFM ground-holding and propagated delay. The taxi-in effect may reflect airfield and gate congestion resulting from large accumulations of ground-held aircraft. The negative en route time effect may result from two factors. First, flights whose routes are not affected by weather may benefit from reduced en route congestion because of ground delays imposed on weather-impacted flights. Second, ground-held flights, once released, may increase flight speeds in order to make up time.

Third, the wind effect is significant and positive for three ASPM delay models, DAFT Total sub-model and En-route sub-model. The IFR effect is significant and positive except DAFT Taxi-in sub-model. The magnitudes of the both effects are, again, fairly consistent across the ASPM and DAFT Total models. Based on the IFR results, an all-IFR day would have delays 20 minutes greater than an all-VFR day, all else equal. The DAFT component models reveal that about 75% of IFR-induced delay is taken at the gate or on taxi-out, and the remaining 25% in the air. The DAFT models also show that wind affects delay primarily by increasing en route flight time. This could mean that TFM responses to wind effects are made tactically, after flights are already in the air. It is also possible that when average wind speeds at the airport surface are high, winds aloft are also high.

¹ The range of scheduled arrivals is [18100, 25400], we multiply the difference by estimated coefficients of scheduled arrivals to calculate the delay difference between low traffic day (18100) and high traffic day (25400).

Finally, the estimation results about weather forecast reveal that over-forecasting increases delay significantly, while the estimated effect of under-forecasting is statistically insignificant, albeit negative. Both results are intuitively reasonable. Forecast convective weather triggers traffic flow initiatives that generate delay even if the weather is not realized. Consistent with this interpretation, the DAFT components that are affected by over-forecasting are those taken before take-off. Realized convective weather, in contrast, will have an effect regardless of whether or not it is forecast. It is interesting that under-forecasting does not appear to increase delays. This suggests that the failure to foresee convective weather affects the cost (and perhaps safety) impacts of delay rather than its quantity.

We notice that in ASPM delay model 3, the coefficient of False-positive is 6.39, while the average False-positive in our sample is 0.40 (refer to Table 1), we estimate the effects of delay by False-positive would be 2.56, which contributes about 20% of the total ASPM delay.²

Of the WITI features, only the mean appears in the parsimonious ASPM Model 3 and DAFT models. Other WITI features are significant in ASPM Models 1 and 2, but became insignificant as other insignificant variables were removed. The message appears to be that other WITI features do matter, but not very much. Specifically, delay increased with WITI dispersion (as captured by the Standard Error), decreases when the WITI distribution has a long right tail (Skewness), increases when storm activity occurs later in the day (Time Mean), and increases when it is more dispersed over the day (Time Standard Error). Further research is needed to confirm these relationships and account for them in operational terms.

4. Conclusions and recommendations

In this paper, we contribute to the growing literature involving the use of econometric models to understand the relationship between observed airline delay and several causal factors, including traffic, airport weather, en route convective weather, and (for the first time, to our knowledge) weather forecast accuracy. Four specific models are presented. Estimation results show the models can explain 60-65% of the observed variation in overall average delay, and 50-60% of the different components contributing to this delay.

² The ASPM delay mean is 13.2 in Table 1.

Many of the results of this analysis are consistent with those of previous research. We emphasize here the findings that are new. First, the use of time indices, such as the DAFT Index discussed here, yield important insights that most conventional average delay metrics do not. The effects of certain factors, such as the amount of flight traffic, are greater when day-to-day variation in flight time (the DAFT total), as opposed to delay (ASPM average delay), is modeled. The combination of schedule buffering and truncating negative delays make standard delay metrics insensitive to much day-to-day variation in operational outcomes. Further, the ability to decompose the DAFT into flight phase components yields important insights about how different causal factors affect delay. To give just one illustration, we found that convective weather generates delay on the ground while slightly reducing average airborne time in the NAS.

Second, our results lend empirical support to the assumption, untested in the previous literature, that an aggregate daily WITI adequately captures the effect of convective weather on delay. Systematic search for other features of the daily time series of WITI values, including those that characterize variation in the magnitude and timing of convective weather impacts, yields some statistically significant results but no dramatic improvements in predictive power.

Third, we find compelling evidence of the effect of convective weather forecast errors on delay. According to our results, over-forecasting is the source of about 20 percent of the total delay in our sample of days. This represents a first-cut estimate of the benefits pool for research and investment intended to improve weather forecasting capability. While our results are less definitive for under-forecasting, they suggest that failure to foresee convective weather does not incur a significant delay penalty. To the contrary, under-forecasting, if anything, reduces delay, although probably at some penalty in terms of cost and safety.

Much remains to be done. The model presented here is first order, excluding the effects of interactions between variables and non-linear effects of individual variables. The scatter plots shown in Section 3 suggest the need for such higher order terms. Additional variables that better capture the effects of terminal delays should also be considered. Further WITI refinements should also be considered. For example, it would be useful to consider weather impacts not just on the weather-free trajectories of flights but also on alternative trajectories that would be used if primary ones are closed due to weather. Finally, there is even more room for improvement on the forecast side, where

the ultimate aim should be to develop a Forecast Impact Traffic Index, or FITI, that reflects not just the aggregate counts of forecast errors but also the locations of these errors relative to flight paths.

References

- [1] Hansen, Mark, and Chieh Yu Hsiao, "Going South? An Econometric Analysis of US Airline Flight Delays from 2000 to 2004", Transportation Research Record, 2005, pp. 1-20.
- [2] Hsiao, Chieh-Yu and Mark Hansen, "An Econometric Analysis of US Airline Flight Delays with Time-of-Day Effects", Transportation Research Record, 2005, pp. 104-112.
- [3] Post, Joseph, Murphy, Daniel, and James Bonn, "A Regression Model of National Airspace System Delay", TRB 86th Annual Meeting, Washington D.C., Jan. 2007.
- [4] Callaham, M., DeArmon, J., Cooper, A., Goodfriend, J., Moch-Mooney, D., and Solomos, G., "Assessing NAS Performance: Normalizing for the Effects of Weather," 4th USA/Europe Air Traffic Management R&D Symposium, Santa Fe, NM, 2001.
- [5] Chatterji, Gano, and Banavar Sridhar, "National Airspace System Delay Estimation Using Weather Weighted Traffic Counts," AIAA Guidance, Navigation, and Control Conference, San Francisco, California, August 2005.
- [6] Sridhar, Banavar, and Sean Swei, "Relationship between Weather, Traffic and Delay Based on Empirical Methods," 6th AIAA Aviation Technology, Integration and Operations Conference (ATIO), Wichita, Kansas, September 2006.
- [7] ASPM online database, <http://www.apo.data.faa.gov/>
- [8] The Real-Time Verification System (RTVS), maintained by FAA Aviation Weather Center, <http://www-ad.fsl.noaa.gov/fvb/rtvs/conv/>

Acknowledgements

The authors would like to thank Dr. Banavar Sridhar and Dr. Sean Swei, at NASA Ames Research Center, for their enthusiastic support and invaluable assistance for this research.

Biography

Mark Hansen is a Professor of Civil and Environmental Engineering at UC Berkeley and co-director of the National Center of Excellence in Aviation Operations Research. He has an undergraduate degree in Physics and Philosophy

from Yale University, and a Masters in City and Regional Planning and a PhD in Engineering Science from UC Berkeley.

Jing Xiong is currently a PhD student at the University of California at Berkeley. She obtained a Bachelor's degree in Structure Engineering at Southeast University in Nanjing, China, and a Master of Science degree in Transportation Engineering at the University of Iowa. Her research interests include air transportation policy and air traffic management.