2.3 QUANTIFICATION OF PREDICTED IMPACT OF WEATHER ON AIR TRAFFIC AND COMPARISON WITH ACTUAL IMPACT

*Alexander Klein**

Air Traffic Analysis, Inc, Fairfax, Virginia

Richard Jehlen

Federal Aviation Administration, Washington, DC

Sadegh Kavoussi, David Hickman, David Simenauer and Mark Phaneuf

AvMet Applications International, LLC, Reston, Virginia

Thomas MacPhail

National Weather Service, Washington, DC

1 Primary Objectives and Activities

Numerous studies, analyses and reports over an extended period have validated the significant impact of weather on the air transportation system – for safety, efficiency and other service expectations (predictability, flexibility, etc). To support performance improvement, rational investment strategies, long-term planning and provide the context for nearer-term performance, standards of measure were needed for both weather and the service expectations to establish a common framework for system stakeholders. In other words, the objective is to establish consistent, quantifiable measures of the present National Airspace System performance linked to weather conditions and project the system into the future to identify areas for "best" redress (investment).

To accomplish the objectives, four general areas of activity have been identified; 1) measure the actual weather impacting the NAS, 2) measure the accuracy of the weather forecast, 3) establish a classification schema for operational strategies, and 4) measure the impact of the strategies in relation to the service expectations.

1.1 Quality (Measure) of Performance of Present System

The first step to satisfy the objectives is the establishment of a consistent, quantifiable measure of weather from the air transportation perspective. That is, a representation weighted to reflect the differing impacts weather imposes on the system as a result of differing demand profiles. For example, the same weather system located between the New York and Chicago metropolitan areas has a greater impact on the air transport system than the equivalent weather system between Chicago and Bismarck, North Dakota.

The National Airspace System (NAS) performance needs to be measured and understood in terms of the "impact – response – outcome" paradigm. Two major factors impacting NAS performance on the "front end" are the inclement weather and traffic demand. Traffic flow management (TFM) units, air traffic control (ATC), airlines and airports respond to these impacts. The different responses to the various weather and demand scenarios result in different outcomes related to the service expectations, such as; Predictability, Flexibility, Access, etc [1]. Specific operational outcomes such as delays, cancellations, excess miles flown, airspace user and service provider costs, can be measured and compared.

The importance of providing objective, consistent measures of front-end impacts on NAS and the resulting performance cannot be overstated. If delays are higher this year than last year, is the increase caused by worse weather, higher traffic demand or the "brittleness" of the current NAS response to disruptions? Or, as a more complex question, if traffic demand is overall similar to last year's but at a number of highly congested airports the demand has in fact increased, and if weather was worse during first

* *Corresponding author address:* Alexander Klein, Air Traffic Analysis, Inc, 3802 Ridgelea Drive, Fairfax, VA 22031

half of the year (compared to last year's) but better thereafter – except perhaps the South – how can it all be related to NAS performance this year vs. last year?

Since weather is a dominant factor affecting air traffic, the two must be considered together when addressing such questions, hence the notion of weather weighted by traffic.

1.2 *"Forecast Weighted by Traffic"*

Following the first step of establishing a consistent framework for measurement of the front-end impacts of weather and demand on the NAS, it is important to recognize the decision loop and cycle-times inherent in the system. Since the air transportation system is comprised of large numbers of independent individuals and organizations operating at national and international scales, decisions are made and logistical commitments entered into hours before the actual event. In the case of a coast-to-coast flight decisions regarding the specific aircraft to be used, route to be flown, fuel needed, etc are made well in advance of the flight itself – which is several hours in duration. So a critical component to system performance is not only what the weather impacting the system is, but even more so what the weather was forecast to be when the critical decisions and commitments were being made.

Again, while accurate weather forecasting is important in every corner of the nation, we should also note that the impact of inclement weather on the national air transportation system is not equal geographically. An area of thunderstorms over the northern Rocky Mountains area will have low impact on the nation's air traffic as a whole (even though impact on local communities can be significant). The same area of thunderstorms over eastern Pennsylvania may cause widespread disruptions to air traffic felt throughout the NAS, due to impact on the extremely busy airspace between Mid-West and New York regions.

So, as in the establishment of the methodology to measure the actual weather, our forecasted metric will also be "weather weighted by traffic", the results of this modified verification will be different, both geographically and in terms of perception. As an example, consider a situation depicted in Fig. 1.

Classic weather forecast verification would conclude that the forecast has missed the weather completely. But from the air transport perspective, the forecast, while somewhat inaccurate, is still valid because the impact of the line of storms on the aircraft is predicted, in the main, correctly.

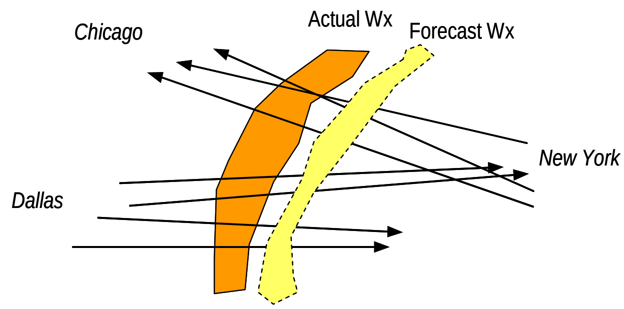


Figure 1: A line of storms impacting a portion of the North-East. Arrows indicate some of the traffic flows

1.3 *Reflect NAS Operational Strategies*

1.3.1 *Types of Strategies*

In response to the multiple variables confronting the system, a range of strategies has been developed to mitigate the impacts in pursuit of meeting the desired service expectations. These impact variables range from the different weather scenarios expected to confront the system (multiplied by the various confidence levels of each forecasted scenario) along with the projections of demand levels (additionally complicated by the different fleet mix within the projection).

Examples of these operational strategies include Ground Delay Programs (GDPs), Airspace Flow Programs (AFPs) and different routing plans, referred to as "Playbook Routes". Each strategy has sub-elements and more detailed choices to fine-tune the impacts to the greatest extent possible. The objective is to impact the fewest number of flights to the least degree possible while ensuring system safety and integrity is maintained.

For this reason, step three is focused on the establishment of a consistent methodology to represent the strategies, their options and any ad hoc responses in a manner that supports comparison. The comparisons would begin to link "strategy suites" (essentially the group of specific strategies, options, etc employed in response to the developing operational picture) to the different weather/demand scenarios.

1.3.2 Performance Outcomes of the Strategies

Considering the wide range of strategies and the permutations that may evolve by employing different portions of each across both the temporal and geographical landscape, the resulting performance outcomes related to the service expectations vary widely. Of particular note, performance management across the NAS covers the broad spectrum of service expectations: safety, security, environment, capacity, flexibility, predictability etc. Many of these areas are mutually interdependent, which means that improving the performance in one area is likely to come at the price of reduced performance in another. This kind of dilemma leads to the need for a "balanced approach" to performance. A "balanced approach" is the result of trade-off decision making between the various performance objectives and targets. For example, implementation of a ground delay program (GDP) will, generally, result in increasing predictability at the possible expense of capacity. As a result, reasonable estimates of the performance outcome are necessary prior to selection and implementation.

Building upon consistent, quantifiable, historical weather representations (and associated forecast accuracies) linked to the specific strategies employed, along with the concomitant outcomes, will support collaboration to weigh the operational strategy candidates within the context of the ultimate performance trade-offs

1.4 Project Future State

While the earlier activities yield a framework used mostly for decision-making on the day of flight based on historical experience, they also provide the mechanisms to evaluate future operating environments. Various "what-ifs" may be proposed and compared to each other through the application of the consistent methodologies and measures guiding the daily operation. For example, future demand projections, including number of flights and fleet mix, can be substituted for the actual, historical demand experience and operational strategies employed. With the relationship to performance across the service expectations in place, estimates of the future NAS level performance will assist in the identification of the "best" redress (investments) to address the shortfalls.

2 Establishing the Weather "Ruler"

2.1 The NAS Wx Index (NWX) Based on Actual Weather

In order to quantify the impact of weather on air traffic, a Weather Impacted Traffic Index (WITI) was proposed by the FAA and the initial version of it was presented in [2]. It was further explored by NASA [3, 4] and FAA researchers [5]. We have expanded and refined these WITI computation methods [6] and have constructed a WITI metric that, at a NAS level, is called the NAS Wx Index (NWX). Key additions to the previously developed WITI metrics are the queuing delay component, the ability to track weather impacts all year round and to drill down to lower levels of detail than previous efforts (individual airports; regions; 1-hour or 15-min time intervals). The basic premise, "weather weighted by traffic", remains unchanged across the range of detail.

NWX measures the <u>front-end</u> impact of weather and traffic demand on the NAS: "the hand that the NAS is dealt every day". Currently the weights of individual components are calibrated to achieve best correlation between NWX and the NAS operational outcome of delays. It must be pointed out that 100% correlation is not a goal in itself: we do expect it to be relatively high and, indeed, observe a correlation coefficient of about 85% between NWX and Delays (with outliers present). At the same time, because we measure only the front-end impact on the system, we use scheduled rather than actually completed number of operations, as well as shortest-path flows rather than actual flight-planned trajectories (the latter include NAS operational strategies to weather – the kind of influence that NWX tries to avoid).

NWX is described in detail in [6]; here, we provide a concise description.

NWX is a composite metric, a weighted sum of three components: En-Route WITI (E-WITI), Terminal WITI (T-WITI), and Queuing Delay (Q-Delay) computed for each major airport every hour. It is aggregated into a NAS-wide number for each 1-hour or 15-min intervals, as well as a daily average, and normalized against a multi-year seasonal average ("ruler", see next sub-section).

NWX components are computed as follows (Figure 2). For en-route component calculation we use "flows" – Great Circle routes between major airports – as ideal, shortest-path unimpeded flight trajectories. To compute the en-route convective weather impact on a flow, we first create a hexagonal grid covering the NAS; typical size of hexagons is ~20 NM. National Convective Weather Data (NCWD) reports are collected in these hexagonal "bins" in 1-hr or 15-min intervals. We then multiply the number of hourly NCWD reports in each hexagonal grid cell that the flow crosses by the hourly frequency of traffic on this flow. The impact is assigned to the two airports that the flow connects – in proportion to the distance of the particular area of weather from the airport: the further away from the airport, the lower the impact. For each airport, the convective impact on all its inbound and outbound flows is thus computed on an hourly (or 15-min) basis.

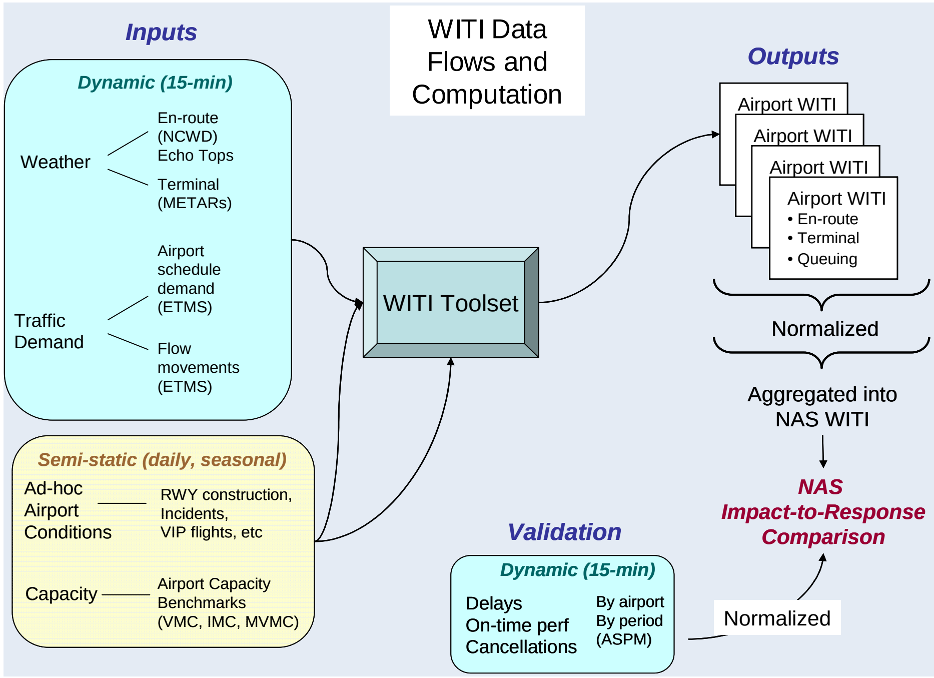


Figure 2: WITI data flows and computation scheme

For terminal weather impact, we use METAR data for major airports (such as those on the OEP-35 list). We maintain a hierarchy of weather factors, from most severe to less severe, so that if, for instance, a thunderstorm was reported and also some rain and wind, the leading factor for the given time interval would be the thunderstorm. To each of these weather factors we assign the corresponding airport capacity degradation percentage: a user-definable parameter. This data is obtained from FAA capacity benchmarks or from historical data analysis. The Terminal WITI is calculated for each airport, every time interval by multiplying the capacity degradation percentage by the number of scheduled hourly operations.

Queuing delay computation is based on evaluating available airport capacity (which can be reduced due to inclement weather) vs. scheduled traffic demand. For each airport, all feasible runway configurations are stored, from best to least-optimal. Each runway configuration comes with its defined arrival and departure capacity benchmarks separate for visual and instrument meteorological conditions (VMC and IMC). Runway configurations are evaluated against current wind velocity, precipitation, cloud ceiling and visibility, as well as maximum allowable cross- and tailwind; from that, the best available runway configuration is found. This also determines the airport's arrival and departure capacity for the 1-hr or 15-min time interval. Additional non-IMC factors such as convective weather in the vicinity of the airport, high winds or heavy snow are considered in terms of impacting airport capacity. From all these factors, we find the one with the most impact and register the corresponding reduced airport capacity for the selected time interval. This is compared to scheduled traffic demand and, if demand exceeds capacity, queuing delays ensue. Delays can accumulate and dissipate as the capacity/demand balance shifts during the day.

2.2 *Establishing the "Ruler"*

In order to compare the NAS performance on different days, months or years, and also to compare the front-end impact on the NAS with system responses and operational outcomes, we need to bring the corresponding metrics to a common denominator, i.e., establish a "ruler". A common method is to normalize NAS metrics against a multi-year average. We chose the 2004-2006 (three-year) NWX average for April-September as the "ruler" whose value is set to 100, and have normalized all NAS Wx Index and airport/regional WITI values against that average. The simple formula is:

$$Normalized\_Index_{daily} = Absolute\_Index_{daily} * 100 / Average\_Index_{seasonal}$$

Importantly, Delay and other NAS operational outcome metrics can be normalized in the same way (2004-2006 April-September average = 100), which allows us to compare impact to outcome directly. In an ideal situation, an average day (in terms of weather and traffic demand; NWX = 100) would result in an average delay (=100 on the normalized scale). If, for example, delay is much higher than NWX for the same day, this may point to system inefficiencies. Delays much lower than NWX typically point to a significant contribution of other factors: for example, during major winter storms the NWX is very high but delays are low due to a large number of cancellations.

2.3 Uses of Current NAS Wx Index

A report including the daily normalized NWX and Delay comparison for the NAS for last the 30 days, as well as monthly NWX and Delay metric averages for the year-to-date vs. two previous years is being provided on a weekly basis for the FAA operational briefing. During convective seasons, the report also includes a chart depicting convective forecast error (presented together with NWX and Delay, Figure 3); this will be discussed in more detail later in this paper.

The NWX report:

- Provides context to the status of the FAA Air Traffic Organization (ATO) progress against performance metrics:
- Identifies areas (anomalies) that may merit further analysis and review of NAS operational performance;
- Is included in the Office of the Administrator (AOA) daily updates, as appropriate, for context regarding ATO performance;
- Is briefed weekly to ATO executives.

The NWX has also been included in the ATO's Chief Operating Officer's presentation to the FAA's Management Advisory Committee (MAC), as well as customer forums, and as part of AOA testimony to Congress.

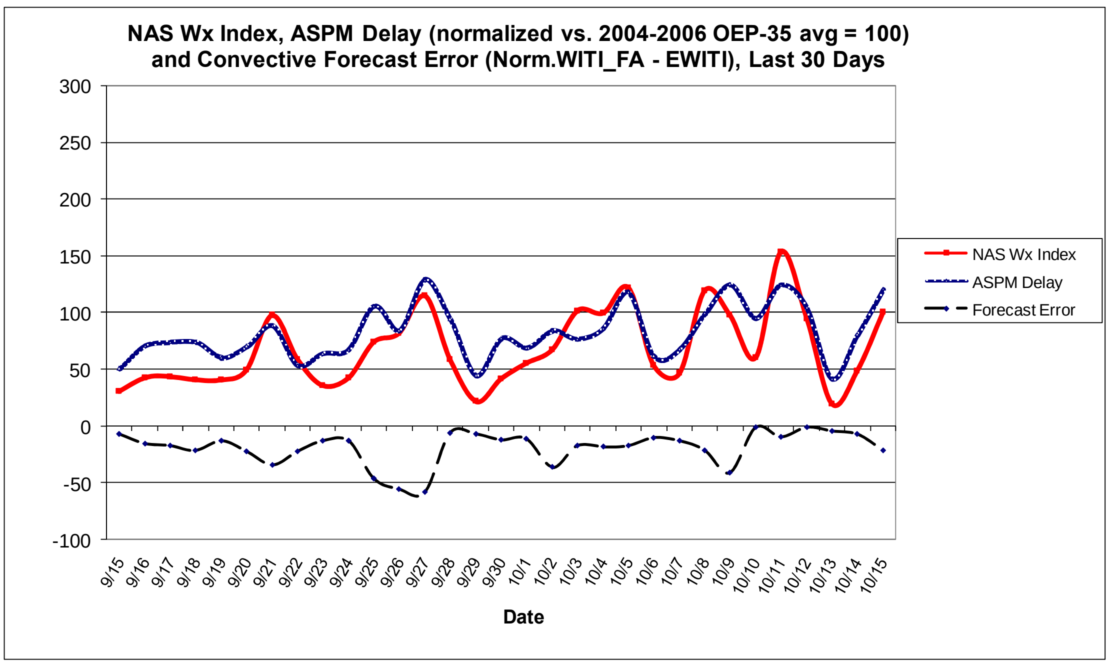


Figure 3: An example of a 30-day normalized NWX, Delay and Convective Forecast Error chart ("ruler" = 100)

3 Creating the NAS Weather Forecast Accuracy Index

3.1 *Objective*

Just as the NAS Wx Index being a variant of Weather Impacted Traffic Index (WITI) metric measures the impact of actual weather on air traffic, its forecast-weather counterpart will measure the forecast weather impact. We call this metric WITI-FA, where FA stands for Forecast Accuracy. The WITI-FA metric is computed using the same scheduled traffic as the WITI metric and both metrics are normalized against the same multi-year WITI average. WITI-FA uses forecast weather (both convective and terminal) while NWX / WITI uses actual weather. In this process, forecast weather data will be converted to "quasi-actual", i.e. will be presented in the same format and style as actual weather data. Since both metrics will be normalized to the same "ruler", we will be able to:

(a) Compare actual and forecast weather impact metrics directly;

(b) Compare both actual and forecast weather impact metrics with the normalized Delay and other operational outcome metrics.

3.2 *Quantifying the Impact of En-Route Convective Forecast on Air Traffic*

The method for quantifying the impact of convective forecast has been presented in [7], which is why, just as with NWX/ WITI computation, we provide only a short description here.

We have focused on the Collaborative Convective Forecast Product (CCFP) at this stage because it is the most widely used and accepted convective forecast product, particularly for en route air traffic flow management. Our goal is to compare the forecast weather impact on traffic with actual impact; to do that we need to convert the forecast convective weather product (CCFP) to actual (NCWD).

CCFP [8], a set of 2-, 4- and 6-hour forecasts, consists of a number of areas; each is characterized by forecast coverage (sparse, medium, solid) and forecast confidence (low, high), see Fig. 4.

The CCFP-to-Quasi-NCWD conversion algorithm can be summarized as follows. We collect hourly NCWD data in hexagonal grid cells covering the NAS. We first compute the maximum possible number of NCWD convective reports, *M,* in a single hexagonal cell in 1-hr period. Then, depending on the coverage and confidence level of a CCFP area that covers this hexagonal cell, we multiply *M* by the two percentages representing the coverage/confidence levels. This yields the quasi-NCWD score for the hexagonal cell derived from CCFP.

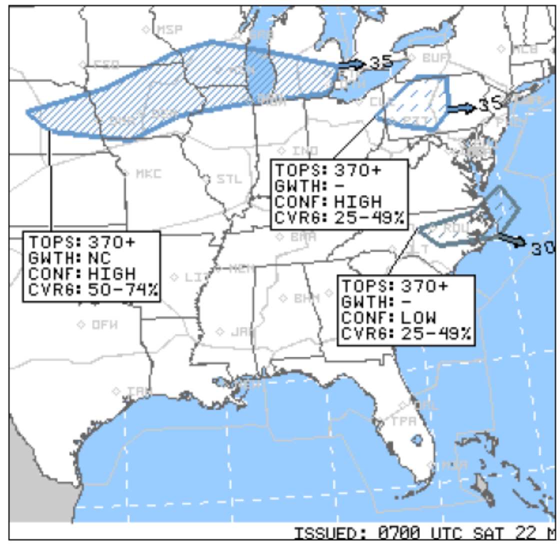


Figure 4: Sample CCFP chart

The NCWD and quasi-NCWD scores for each hexagonal cell are used for En-Route WITI (E-WITI) computation; see sub-section 2.1. Further specifics of quasi-NCWD and convective WITI-FA computation, such as the use of 1-hour intervals and the interpretation of CCFP coverage and confidence intervals, are presented in [9].

The next series of pictures (Fig. 5) illustrates how the verification of convective impact forecast may be different from verification of convective weather forecast. The situation depicted on all three parts of Fig. 5 is from one of the worst convective-impact days of 2007, June 8th, at 2200Z hours.

The first picture (Fig. 5, top) shows NCWD coverage (top left), quasi-NCWD coverage (top right) and their difference, or "Delta". Hexagonal cells where Quasi-NCWD score was higher than NCWD are depicted in pink (overforecast); cells where Quasi-NCWD was lower are shown in blue (underforecast). Both NCWD and Quasi-NCWD is depicted in red; the intensity the red color reflects the intensity of convective weather (number of convective reports in hexagonal cells in 1-hr intervals). Overall, convective weather seems to have been overforecast.

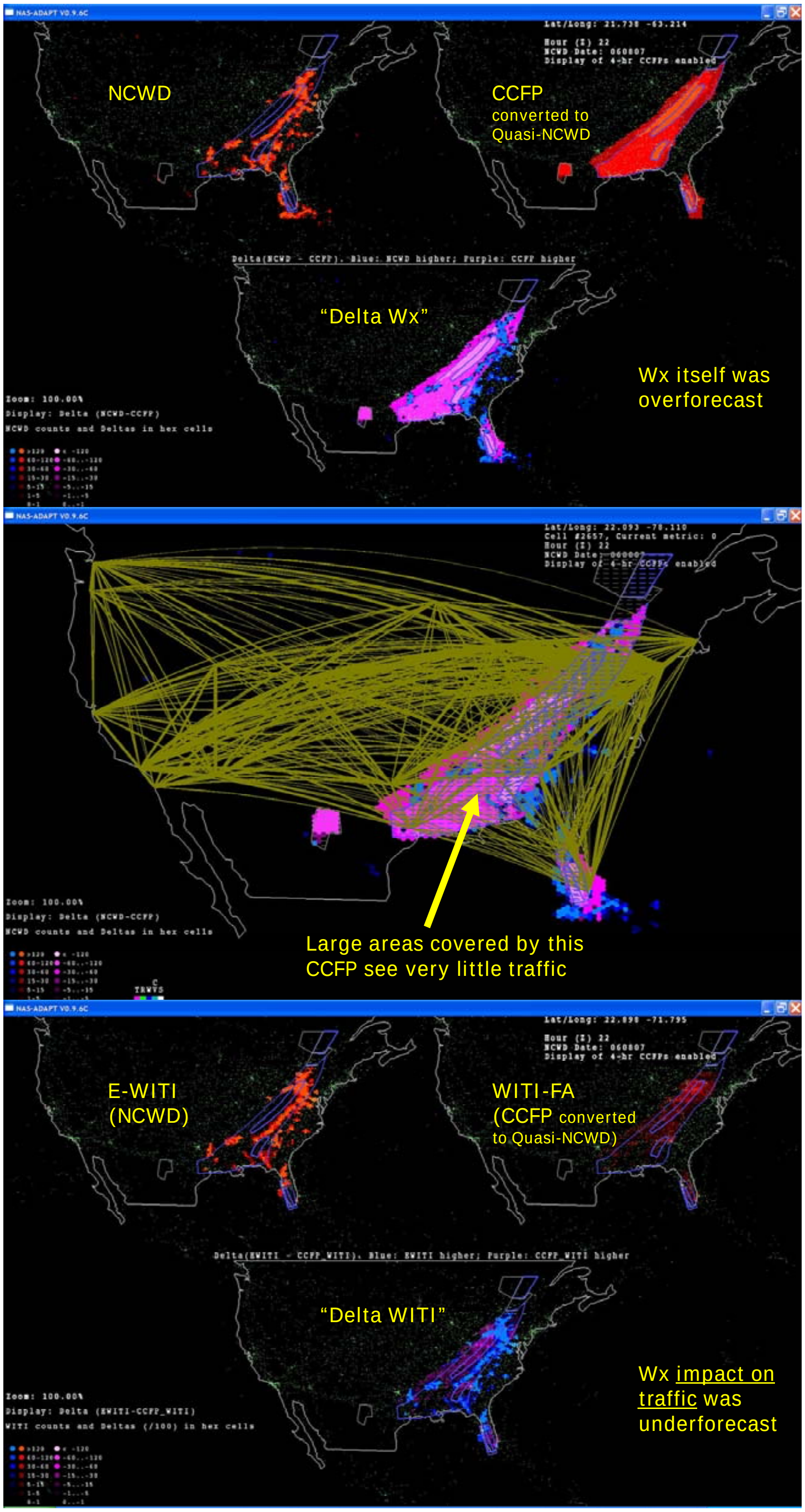


Figure 5: Forecast convective impact on air traffic vs. convective weather forecast (explanation in text)

The second picture (Fig. 5, middle) shows the main flows across the NAS; the flows connect major airports belonging to the OEP-35 group (a larger list of airports, such as OEP-55, and the flow connecting them, can be used if desired). One can see that large areas of forecast weather do not see much traffic scheduled to fly through them.

The third picture (Fig. 5, bottom) shows a different metric in hexagonal cells: E-WITI and E-WITI-FA instead of NCWD and quasi-NCWD-derived-from-CCFP, respectively. Once the "weather weighted by traffic" principle has been applied, the situation reverses: convective weather impact has been underforecast.

Correlation between WITI and the three WITI metrics derived from 2-, 4- and 6-hr CCFP for the same 30-day period as shown in Figure 3 is quite high, as the next chart illustrates (Fig. 6), with a general tendency to underforecast.

This tendency can be traced back to the fact that most of CCFP areas are low-confidence, sparse-coverage or high-confidence, sparse coverage. Even though CCFPs cover large geographical areas, their "convective density", so to speak, is relatively low compared with, say, the density of lines of storms sweeping through the Eastern half of the United States during the convective season. The impact of such lines of storms on air traffic can be quite high. The E-WITI and E-WITI-FA metrics capture these differences in "convective density".

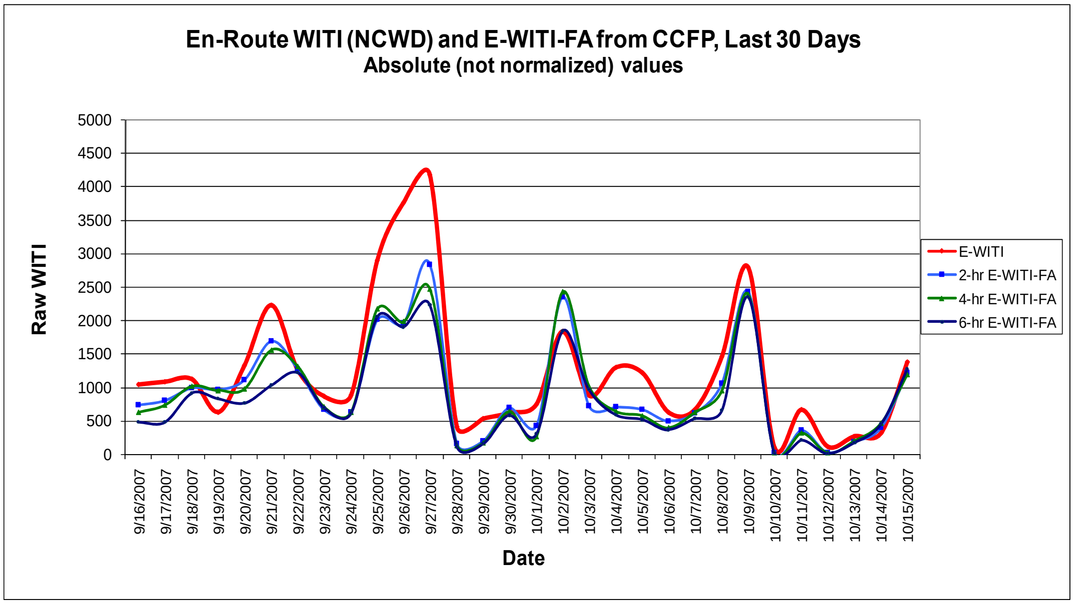


Figure 6: E-WITI and E-WITI-FA

3.3 Terminal Weather Forecast

Just as the E-WITI-FA metric was constructed for the en-route convective weather forecast, the T-WITI-FA metric can be created to quantify the forecast terminal weather impact on air traffic (in this case, operations at each airport). For the actual weather, we use METAR data; for the forecast weather, a natural choice is the Terminal Area Forecast (TAF). Since this has not been reported previously, we will devote somewhat more space to this methodology here.

The first task is to convert a TAF dataset for a day (sample shown in Fig. 7) to METAR format (Fig. 8).

```
FTUS43 KLOT 302300
KORD 010423Z 010424 05012KT P6SM BKN250
   FM0900 12007KT P6SM BKN250
   FM1400 24012KT P6SM VCTS BKN035CB
   FM1800 30012G18KT P6SM BKN040
   FM2200 01012KT P6SM FEW050=
```

Figure 7: TAF sample

metar	KORD	2006	4	9	1456	null	999	10	10	7	-3	null	-2	8	49	-1	4	30.29	null
metar	KORD	2006	4	9	1556	null	999	10	10	9	-3	null	0	9	43	250	5	30.29	BR
metar	KORD	2006	4	9	1656	null	250	10	10	11	-3	null	1	10	38	120	5	30.27	RABR
metar	KORD	2006	4	9	1756	null	250	10	10	11	-3	null	1	10	38	0	0	30.24	null
metar_sat	KORD	2006	4	9	1756	null	250	10	10	11	-3	null	1	10	38	0	0	30.24	null

Figure 8: METAR sample

Probabilistic elements of a TAF dataset, namely the TEMPO and PROB data blocks, need to be converted to a deterministic (METAR-like) format.

In this process, the TEMPO group is treated as follows. Since a TEMPO forecast is predicted to be valid for at least half of the specified time interval, we treat it as having approximately 50% probability. From our analysis of TFM and pilot decision making, significant weather with 50% probability will be treated as 'definitely not to be ignored'. Therefore, we treat TEMPO forecasts as having 100% probability for the time interval specified. The PROB group is treated differently; since the probability is usually 30% (i.e., less than 50%), we "roll the dice" and assign PROB weather events to roughly 1/3 of the hours specified in the PROB time span.

The software scans the TAF dataset and constructs series of complete daily METAR-like datasets for each airport. The datasets are:

- The so-called "Best Fit". For this dataset, we use all the TAFs and amendments issued during the day; the end result should ideally match the METAR for the same day / airport. In reality, the two datasets are close but not always identical.
- 2- and 4-hour "look-ahead TAFs". Here, we process an entire set of TAFs and amendments throughout the day but, when reading a TAF or TEMPO group issued at, say, 0600Z, we block out the first

2 or 4 hours of this TAF or TEMPO, respectively.

- "Pure" TAFs issued at 0000Z, 0600Z, 1200Z and 1800Z. No amendments are processed.

Each dataset creates a forecast "stream" of weather events. As described in sub-section 2.1, each hourly observation (or in this case, forecast) leads to airport capacity degradation if inclement weather was observed (forecast, respectively). The T-WITI-FA metric is then computed as the forecast percent capacity degradation multiplied by the number of scheduled hourly operations at the airport. Next, T-WITI-FA metrics are generated based on the above mentioned TAF datasets; they can be compared with the T-WITI generated using actual (METAR) weather observations.

A convenient way for comparing WITI and WITI-FA metrics, first proposed in [7] and used for convective weather impact illustration, is to construct the so-called NAS-Day-at-a-Glance Matrices. As an example, the T-WITI matrix for May 16, 2007 is shown in Fig. 9.

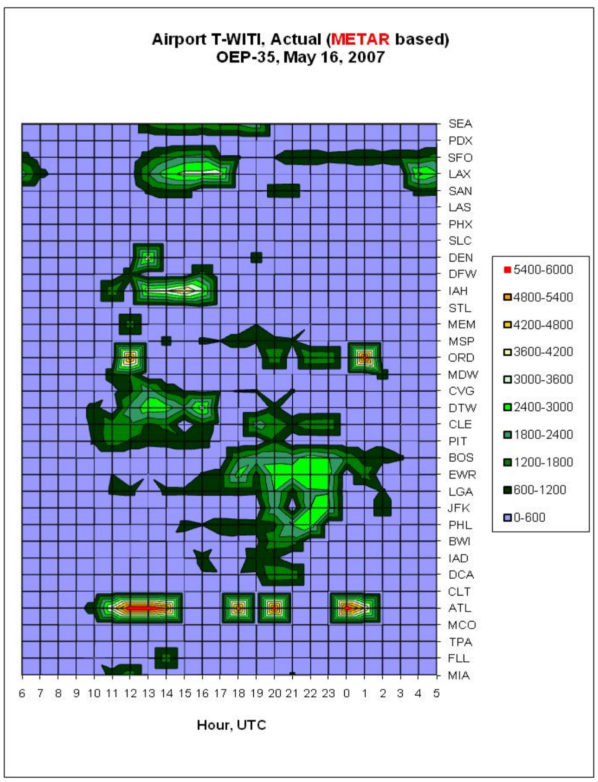


Figure 9: T-WITI NAS-at-a-Glance matrix

Each such matrix has the OEP-35 airports shown in geographical order (coast-to-coast) on the vertical axis and the hours of the day on the horizontal axis. The matrix thus shows weather *impact* (not just weather itself) on traffic to/from the nation's major airports.

The T-WITI and T-WITI-FA scores in airport/hour boxes have been normalized vs. 2004-2006 April-September NAS average, which would be 100 on this scale. Thus, areas of red (hourly T-WITI for an airport much higher than 100) mean that terminal weather impact for this airport at this hour was much higher than the 3-year NAS average. Areas of blue indicate airports/hours with below-average terminal weather impact.

In the example in Fig.8, there was significant terminal weather impact in the North-East (due to low ceilings/visibility) and in Atlanta (local thunderstorms and IMC). This can be compared with a "Best-fit" T-WITI-FA matrix for the same day that uses TAF instead of METAR (Fig. 10).

Comparison of Fig. 9 and 10 shows that the forecast and actual terminal weather impacts are similar but not identical, even in the "Best Fit" case. Initial analysis of reasons for such discrepancies shows that wind gusts are often overforecast and the information is not always amended.

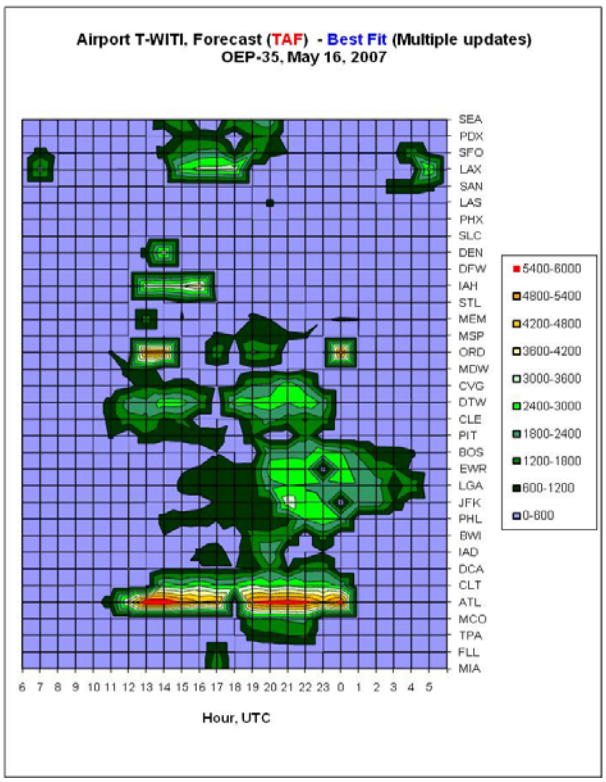
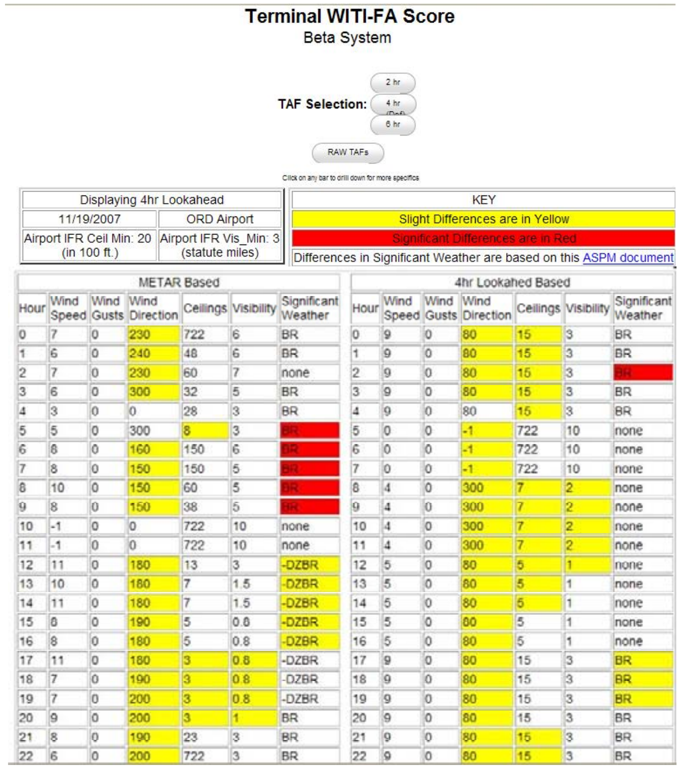


Figure 10: T-WITI-FA NAS-at-a-Glance matrix

For example, we have compared the forecast and actual winds at Philadelphia on May 6, 2007. The wind speed forecast was quite accurate. Wind gust forecast started out as accurate but from



Terminal WITI-FA Score
Beta System

TAF Selection: 2 hr / 4 hr (Def) / 6 hr

RAW TAFs

Click on any bar to drill down for more specifics

Displaying 4hr Lookahead	
11/19/2007	ORD Airport
Airport IFR Ceil Min: 20 (in 100 ft.)	Airport IFR Vis_Min: 3 (statute miles)

KEY
Slight Differences are in Yellow
Significant Differences are in Red
Differences in Significant Weather are based on this ASPM document

METAR Based							4hr Lookahed Based						
Hour	Wind Speed	Wind Gusts	Wind Direction	Ceilings	Visibility	Significant Weather	Hour	Wind Speed	Wind Gusts	Wind Direction	Ceilings	Visibility	Significant Weather
0	7	0	230	722	6	BR	0	9	0	80	15	3	BR
1	6	0	240	48	6	BR	1	9	0	80	15	3	BR
2	7	0	230	60	7	none	2	9	0	80	15	3	BR
3	6	0	300	32	5	BR	3	9	0	80	15	3	BR
4	3	0	0	28	3	BR	4	9	0	80	15	3	BR
5	5	0	300	8	3	BR	5	0	0	-1	722	10	none
6	8	0	160	150	6	BR	6	0	0	-1	722	10	none
7	8	0	150	150	5	BR	7	0	0	-1	722	10	none
8	10	0	150	60	5	BR	8	4	0	300	7	2	none
9	8	0	150	38	5	BR	9	4	0	300	7	2	none
10	-1	0	0	722	10	none	10	4	0	300	7	2	none
11	-1	0	0	722	10	none	11	4	0	300	7	2	none
12	11	0	180	13	3	-DZBR	12	5	0	80	5	1	none
13	10	0	180	7	1.5	-DZBR	13	5	0	80	5	1	none
14	11	0	180	7	1.5	-DZBR	14	5	0	80	5	1	none
15	8	0	190	5	0.8	-DZBR	15	5	0	80	5	1	none
16	8	0	180	5	0.8	-DZBR	16	5	0	80	5	1	none
17	11	0	180	3	0.8	-DZBR	17	9	0	80	15	3	BR
18	7	0	190	3	0.8	-DZBR	18	9	0	80	15	3	BR
19	7	0	200	3	0.8	-DZBR	19	9	0	80	15	3	BR
20	9	0	200	3	1	BR	20	9	0	80	15	3	BR
21	8	0	190	23	3	BR	21	9	0	80	15	3	BR
22	6	0	200	722	3	BR	22	9	0	80	15	3	BR

Figure 11: Comparison of METAR and "4-hr look-ahead" TAF data for ORD airport

1400Z, gusts exceeding 30Kt were forecast whereas actual gusts did not exceed 20 Kt. The difference is significant because winds or gusts in excess of 30 Kt are classified as high winds in WITI and the capacity degradation percentage for the airport is higher than for winds/gusts below 30 Kt. As a consequence, T-WITI-FA for Philadelphia would show overforecast for May 6.

Differences between forecast and actual impact at individual airports can sometimes be significant. To help develop a better understanding of these differences and their effect, we have developed a methodology that allows the analyst to compare METAR and TAF for any airport / day from the impact-on-aviation standpoint.

Figure 11 shows an example of the implementation of this methodology. METAR and TAF for Chicago O'Hare airport (ORD) are shown for November 19, 2007. Wind speed, gusts, direction, cloud ceilings, visibility and significant weather (if any) are shown, as read from METAR (Fig. 11, left) and TAF (Fig. 11, right) data and interpreted by WITI software. Here, we use a "4-hour look-ahead TAF": that is, for any given hour N we consider only the TAFs that refer to hours N+4, N+5 and further. This corresponds to a 4-hour tactical planning horizon and is correlated with the 2-, 4- and 6-hour look-ahead CCFPs discussed earlier in this paper. The "look-ahead" TAF is compiled from the TAFs issued continuously throughout the day.

Differences between corresponding METAR and TAF table cells are highlighted in yellow if they are sufficient to cause a different degree of potential airport capacity degradation. For example, METAR ceilings for the early portion of the day are above ORD VFR minima while the TAF forecast ceilings are below (i.e. visual approaches are not possible); the corresponding TAF table cells are highlighted in yellow. If the difference is even more significant, such table cells

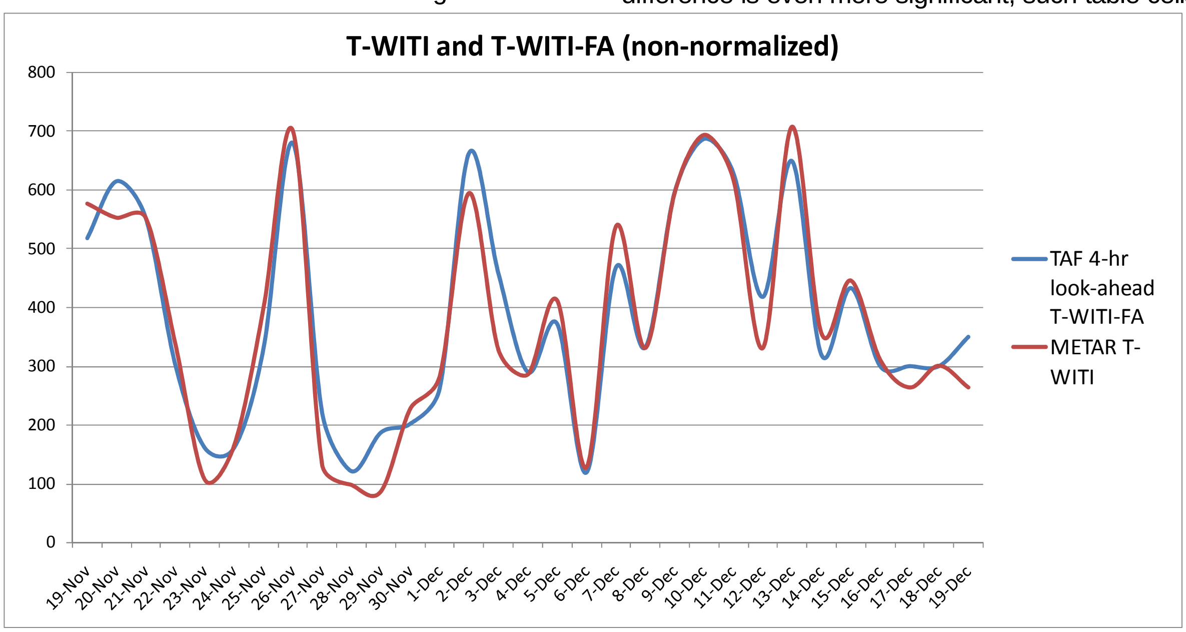


Figure 12: T-WITI and a 4-hour TAF look-ahead T-WITI-FA, Nov-Dec 2007

are shown in red. For example, "BR" (rain/mist) in METAR section vs. no significant weather in the other section.

A sample 30-day comparison chart for T-WITI/T-WITI-FA is shown in Fig. 12.

Key trends characterizing these discrepancies and the possible explanations will be explored in the course of further research.

3.4 Relationship between Convective Forecast Error and Delay

Looking at the chart showing NAS Wx Index, Delay and Convective Forecast Accuracy Error, such as that shown in Fig. 3, one can see that the magnitude of the Forecast Accuracy Error is proportional to the Delay and NAS Wx Index metrics. But can we rephrase this and say that delays are proportional to the forecast accuracy error? This would stop short of an even stronger statement that weather forecast accuracy is a major factor contributing to NAS delays.

A scatter plot depicting correlation between delays and convective forecast accuracy ("E-WITI minus E-WITI-FA Delta") for three convective seasons, 2005, 2006 and 2007, is shown in Figure 13.

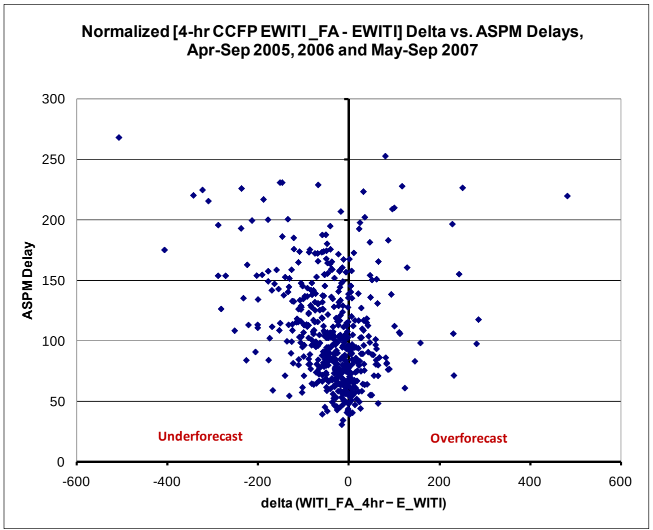


Figure 13: Delta E-WITI-FA vs. ASPM Delay

Here, we distinguish between overforecast (positive Delta) and underforecast (negative Delta). The correlation between the two variables is visible but is more significant for the underforecast cases.

However, correlation alone does not mean that there is a strong causality link between large convective forecast errors and high delays. Both are observed on high weather impact days. Moreover, there are days with high delays but relatively low convective forecast error (dots close to vertical axis with values exceeding 150 in Fig. 13). In order to establish how strong this causality may be, we will need to investigate NAS response strategies and operational outcomes. This next phase of our research is pending; its outline is provided below.

4 NAS Operational Strategies and Operational Outcomes vs. the Accuracy of Weather Impact Forecasting

The first activity to establish a framework for stakeholders regarding the weather was the establishment of standards for measurement and comparison. The same is true for developing the framework for the operational strategies selected to mitigate the impact of weather and the resultant operational outcomes from their application.

4.1 Methodology to Represent ATM Operational Strategies

When confronted with a given (or forecast) weather phenomenon, there are a suite of known "mechanisms" or strategies to mitigate the impact at a system level. These strategies include Ground Delay Programs (GDPs), Airspace Flow Programs (AFPs), routings (including "Playbook Routes") and Miles-in-Trail (MIT). Each of these may be employed independently, in conjunction with the others, applied in some locations, but not others and at varying levels of implementation. Each mechanism, or combination of mechanisms, results in differing impacts across the service expectations mentioned earlier.

Discussion of the relative merits of one strategy vice another requires the establishment of a consistent, quantifiable representation of the strategies employed. The objective is to represent the responses in a structured manner so they can be "mapped" against the forecast, the actual weather and the ATM System performance from the perspective of the service expectations.

4.2 Description of ATM Operational Strategies

Below is a brief description of the ATM operational strategies.

- Ground Delay Program (GDP) is a mechanism focused on a specific airport. Under it, arrival slots are allocated to flights scheduled to land within the period of the GDP. The flights eligible for the GDP may be limited by time, distance from the object airport or apply to only a single operator.
- Airspace Flow Programs are similar to GDPs in that they allocate "slots" through identified volumes of airspace. Again, specific inclusion may be applied based on a group of characteristics (time, etc).
- Routings are used to identify specific routes that are published to avoid weather. These may be limited to a single small route segment or involve a coast-to-coast route comprised of multiple segments for large weather systems. Playbook routes refer to a collection of pre-defined, published routes and responses that support communication of complex strategies to ensure clarity and effective actions within the decision or implementation cycles.
- Miles-in-Trail refers to the requirement that successive aircraft satisfy a specific longitudinal separation for flow planning purposes.

4.3 Compare ATM Strategies Employed

Once established, the consistent, quantifiable (normalized) records of the weather, both forecast and actual, the ATM operational strategies employed, and the resultant service effects may be "mapped" against each other. The results of the analysis will identify the "best" strategies, based on the historical record, to shape a certain outcome within the context of the service expectations.

4.4 When Forecast Is Perfect but Weather Impact Is Still Significant

Currently, when ATM performance is measured following a weather event there are significant uncertainties surrounding the following questions:

- How accurate was the forecasted weather?
- How accurate was the traffic demand?
- If the forecast and traffic demand were "perfect", was the best ATM operational strategy for that scenario selected and what impact did the ability to execute this strategy have on ATM System performance? (That is, "Was the correct strategy poorly executed?")

The answers to these questions are the drivers across the spectrum of activities to develop consistent, quantifiable methodologies to measure and describe the weather, as it was encountered and forecast, ATM operational strategies and service expectations.

4.5 Using the Forecast Accuracy Metrics to Identify Areas for Improvement

One of the first benefits of the NAS Weather Index and the forecast accuracy based upon it has been the ability to provide a common, consistent lexicon of performance between weather specialists and air traffic management specialists. It has begun to inform the discussion regarding what products are best suited to positively impact ATM System performance.

5 CONCLUSIONS

We have established a framework for consistent, objective measurement of the "front-end" impact of weather and traffic demand on the NAS (WITI/NWX), the forecast accuracy impact (WITI-FA), and the operational outcomes (delays, costs). We have also set the objectives for quantifying the impact of ATM operational strategies that are implemented in response to weather and traffic demand. The methodology and metrics being developed will allow us to "connect the dots" between the impact, the forecast accuracy, the response strategies and the operational outcomes through a set of NAS performance indicators that are normalized using the same "ruler". This will help improve collaborative ATM operation and planning, as well as development of aviation weather forecast products and ATM decision support tools.

REFERENCES

[1] International Civil Aviation Organization, ICAO, Document 9854 (AN/458), *Global Air Traffic Management Operational Concept*, Appendix D, First Edition – 2005.

[2] Callaham, M., J. DeArmon, A. Cooper, J. Goodfriend, D. Moch-Mooney, and G. Solomos, *Assessing NAS Performance: Normalizing for the Effects of Weather*, 4th USA/Europe ATM R&D Seminar, Santa Fe, NM, 2001.

[3] Chatterji, G., and Sridhar, B., *National Airspace System Delay Estimation Using Weather*

Weighted Traffic Counts, AIAA Guidance, Navigation, and Control Conference, San Francisco, CA, 2005.

[4] Sridhar, B., *Relationship Between Weather, Traffic, and Delay based on Empirical Methods,* NEXTOR NAS Performance Workshop, Asilomar, CA, 2006.

[5] J. Post et al, *The Use of Flight Track and Convective Weather Densities for National Airspace System Efficiency Analysis,* 21st DASC, Piscataway, NY, 2002.

[6] Klein, A., R. Jehlen, and D. Liang, *Weather Index with Queuing Component for national airspace system Performance Assessment* . 7th USA/Europe ATM R&D Seminar, Barcelona, Spain, 2007.

[7] Klein, A., Day's Weather In the NAS: Visualization Of Impact, Quantification And Comparative Analysis. *Proceedings of the 6th NASA ICNS Conference, Baltimore, MD, 2006.*

[8] Collaborative Convective Forecast Product: Product Description Document. The US NOAA / National Weather Service, at: http://aviationweather.gov/products/ccfp/docs/pdd-ccfp.pdf

[9] Klein, A., S. Kavoussi, D. Simenauer, D. Hickman, M. Phaneuf, and T. MacPhail, *Using a Convective Weather Forecast Product to Predict Weather Impact on Air Traffic: Methodology and Comparison with Actual Data*, 7th NASA ICNS Conference, Herndon, VA, 2007.

ACKNOWLEDGEMENTS

This project was funded by the National Weather Service in coordination with the FAA.

The authors are grateful to Kevin Johnston of NOAA/NWS for his support and guidance.