DO FLIGHT TIMES CHANGE YEAR TO YEAR? A COMPARISON OF 2001 AND 2002

G. Solomos[¥], D. Knorr[‡], M. Blucher [¥], J. DeArmon [¥], G. Dorfman [¥], H. Yi [¥]

[¥]The MITRE Corporation
Center for Advanced Aviation System Development
7515 Colshire Dr., McLean VA 22102

[‡]Federal Aviation Administration
Free Flight Program Office
1500 K Street, NW, Washington, DC 20005

Abstract

For a system to be managed, it must be measured. The National Airspace System (NAS), the collection of airspace and airport resources in and under the control of the United States (U.S.), is a very complex system which presents great challenges of measurement and management. One key measure of en route airspace efficiency is flying time, i.e., the amount of time it takes for an aircraft to travel through en route airspace on its flight from origin to destination airport. A number of factors influence flying time, the most obvious being the winds. Other important factors are: traffic congestion, air traffic management (ATM) interventions, route structure, industry strategies, and weather.

In this paper, we calculate en route flying time in the aggregate and compare a subset of data for the years 2001 and 2002. We account for aircraft equipment type and adjusted flying time for wind effects. In addition, we have selected for analysis a sample of only good weather days (15 in each of the 2 years).

The results show that 2002 has slightly shorter flying times than 2001, on the order of 20–25 seconds shorter. Although we did not investigate the causes for the change, we conjecture that the lower levels of traffic after the Fall of 2001 are causing less congestion and less delay. New automation and procedural initiatives may also be contributing to the improvement.

Introduction

The Federal Aviation Administration (FAA) is charged with the safe, efficient movement of air traffic in U.S. airspace. In recent years, the FAA has begun initiatives to measure efficiency[*] of airspace usage. The measurements are useful for determining the impact of automation and procedural enhancements. Such measurements are also useful for identifying problem areas in the NAS, which can be targeted for remediation.

Changes in efficiency may be examined as a function of time. This study assessed system efficiency using en route flying times, comparing 2001 versus 2002. Although several studies have been performed in the past calculating changes in en route flying distance, the authors are familiar with one time-based analysis in the open literature. Bolczak, et al.[1] used Estimated Time of Arrival (ETA) data to analyze a trend in flying time, and discover some year-to-year changes.

Other studies of flight times are in the literature, though they have not compared actual or modeled values across years. An early simulation of Free Flight,[†] Ball, et al.[2] analyzed flying times for flights without the constraints of route structuring.

---

[*]Efficiency may be defined as the level of utilization of a resource, with consideration of the cost or effort undertaken to achieve that level.

[†] Free Flight is an industry/government initiative which provides greater freedom for pilots and airlines to select planned and flown routes and take-off times. Free Flight supports collaboration between airspace managers and airspace users.

Willemain[3] examined sources of variability in flying times for certain city pairs. The Bureau of Transportation Statistics[4] has collected flight time statistics and hosts a web site which allows the public to access this information.

In this study, results are presented in aggregated form, with some detail with respect to city pairs. Differences were found in the flying time metrics: 2002 has slightly lower flying times than 2001. The causes for this difference are not discussed.

Scope

This study calculated three metrics for en route flying time in the NAS: total flying time, minimum flying time, and excess flying time (with respect to the minimum). Comparing year-to-year flying time using three metrics gives a more complete characterization than using a single metric. Note that although these metrics are measured in the en route domain, they need to be studied in relation to other metrics. By themselves, they are not a sufficient indicator of en route efficiency. For example, changes in the en route environment often manifest as changes in ground delay.

An important adjunct calculation was undertaken in computing the flying time metrics, that being the effect of winds. Flying times are approximated using actual, observed times of flights in the air, and then adjusted by a pre-computed "wind effect," as will be explained in the section "Adjusting for Winds."

This initial study focused on a sample of good-weather days in the NAS. In the future, a bad weather day will also be studied and then compared to these preliminary findings.

Approach

This study used data in the time period of January to August for the years 2001 and 2002. We avoided the September–December time frame in 2001 due to the large, unplanned schedule changes.

We selected 15 good-weather days from each of 2001 and 2002 as a population of 30 days for study. Days were selected not based on meteorological reports, but rather, using a scoring scheme. This scoring scheme ranked the days of the year from best to worst with respect to a composite measure which considered percent of flight cancellations, percent of flight diversions (landing at an airport other than the one scheduled), and percent of flights with more than 30 minutes of departure delay. These measures were based on Airline Service Quality Performance (ASQP)[5] data. Therefore, some less-than-ideal weather may exist in some places on a given day. However, the impact of that weather was minimal compared to other days in each year. Previous studies such as Callaham[6] showed a good correlation between this flight performance measure and general weather conditions in the NAS.

To capture en route flying times, we analyzed flying times not from airport to airport, but from 100 nautical miles (nmi) from origin airport to 100 nmi from destination airport. By setting these bounds, we have, to a large degree, removed terminal area constraints such as delay maneuvers and slow downs. The data source was Enhanced Traffic Management System (ETMS) data.[‡] For simplicity, we required that both origin and destination airports be located in the conterminous U.S. (CONUS).

Adjusting for Winds

Winds, especially winds aloft which are high velocity and impact flights at cruise altitude, have a major impact on the flying time of a flight. In general, in the CONUS there are so-called "prevailing westerlies." That is, the main wind flow is from the west to the east, although there are significant local deviations from this trend on some days.

To adjust for winds, it would not be accurate to use wind speed and direction data and simply apply vector algebra to adjust flying times. This is because of pilots' compensating actions—faced with strong headwinds, pilots may "throttle

[‡] ETMS data is collected from about 40 air traffic control (ATC) computer systems around the country. Flight data, including departure, arrival, en route positions, and flight plan information, are collected, processed, and disseminated to Traffic Flow Management (TFM) facilities and other customers. Oiesen[7] provides a good overview.

forward"; faced with strong tailwinds, pilots may "throttle back."

Our approach for computing wind effects was to analyze "opposing" traffic—traffic in opposite directions at the same time will tend to experience opposite winds. For each aircraft equipment type (as specified in ETMS), we calculated average speed (ground flight track distance divided by time) for flights from airport A to airport B, and also for flights from airport B to airport A. It was assumed that, if winds were not a factor, traffic in each direction (for a given aircraft equipment type) would want to travel at about the same speed. (Using this assumption, the calculation would pick-up another important effect besides winds: flow control impedance. This effect needs to be assessed in the context of this study, and will be pursued in ongoing research.) It was assumed half the difference between the two speeds was the wind effect. For example, if the average speed from A to B was 400 knots (kts), and the average speed from B to A was 600 kts, then the wind effect was 100 kts (= (600-400)/2).

This computed wind effect was applied to each flight to obtain an adjusted speed. A wind effect adjustment was computed and applied for each day, for each origin/destination group (see section titled Grouping Airports) and each equipment type.

For example, on a given day, assume an individual flight from A to B had a speed of 410 kts, and a flight (of the same aircraft equipment type) from B to A had a speed of 588 kts, and the wind effect (between A and B for that aircraft equipment type) was 100 kts, per the above example. The A-to-B flight would get an adjusted speed of 510 kts (410 kts observed + 100 kts wind effect), and the B-to-A flight would get an adjusted speed of 488 kts (588 kts observed – 100 kts wind effect). Using these adjusted speeds, an adjusted *flying time* was calculated for each flight on a daily basis: adjusted flying time = track distance flown / adjusted speed.

Comparing Unadjusted and Adjusted Flying Times

To show the effect of applying wind adjustments to individual flight to adjust their flying times, see Figures 1 and 2. For both figures, the x-axis is the flying time, labeled *Travel Time* (for flights to/from Los Angeles International [LAX]/Washington Dulles International [IAD]), and the y-axis is the number of flights labeled *Frequency*. Figure 1 shows two distributions for unadjusted flying times: on the left are the faster, eastbound flights; on the right are the slower, westbound flights. Figure 2 (which has a different vertical axis scale than Figure 1) shows these same two distributions with wind adjustment. Note that the distributions in Figure 2 are nearly coincident in shape and position on the x-axis, in contrast to those in Figure 1. The height gap between the distributions in Figure 2 is caused in part by the difference in flight counts—there are about 30 more flights westbound.

The merging of two distinct curves in Figure 1 to the nearly coincident location and shapes of curves of Figure 2 indicates that the adjustment for winds is successful. The variation caused by wind direction has been largely eliminated, evidenced by the convergent shifts. In addition, the variation caused by aircraft equipment type, shown as the sinuosity of the Figure 1 curves, has been removed in Figure 2.

Grouping Airports

As noted, winds differ from day to day, and hence adjustments need to be computed and applied on a daily basis. In order to apply the "opposing traffic" methodology, sufficient numbers of flights for each aircraft equipment type need to exist between origin/destination pairs for meaningful averages.

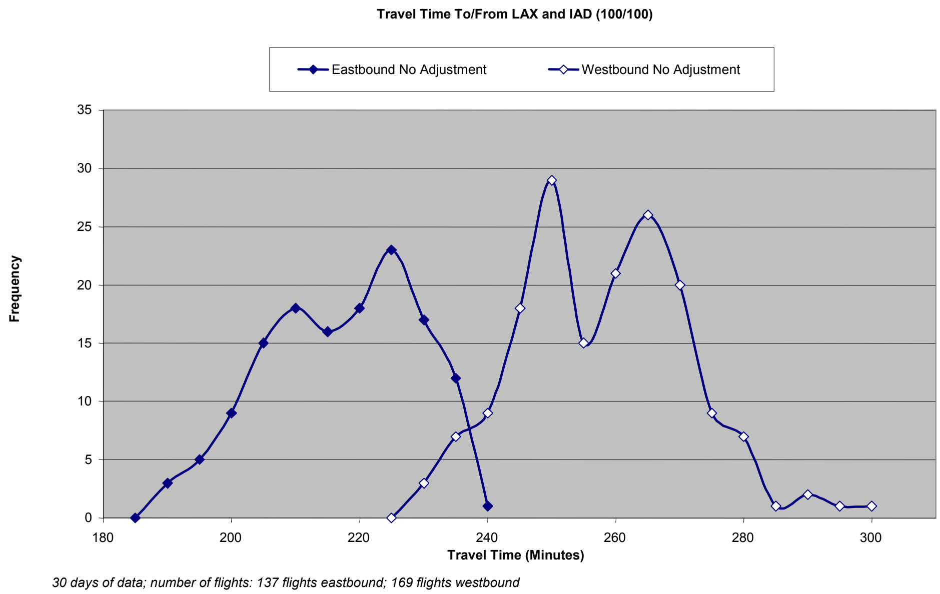

Figure 1. Comparing Flying Times, for LAX to/from IAD (No Wind Adjustment)

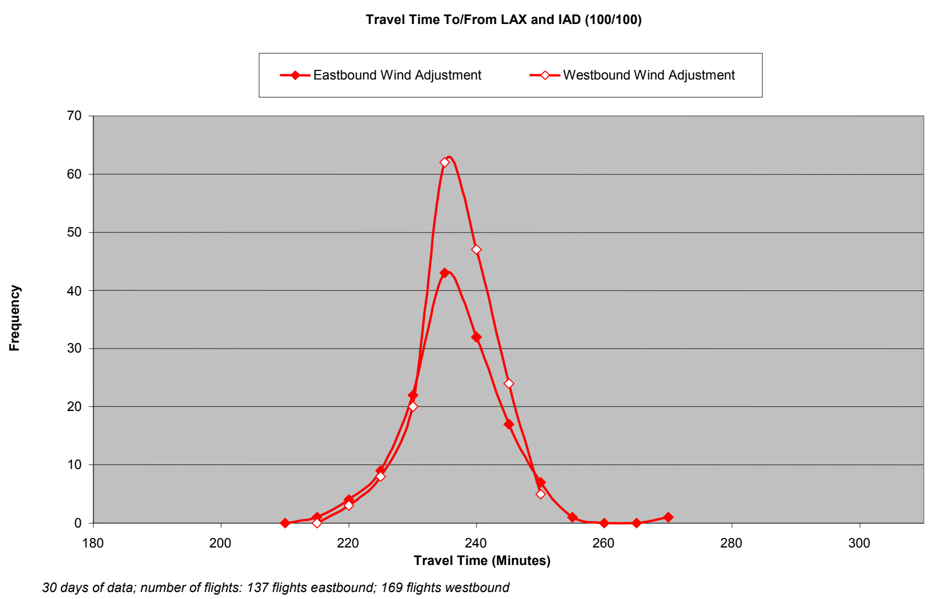

Figure 2. Comparing Flying Times, for LAX to/from IAD (With Wind Adjustment)

In order to calculate a wind adjustment value, we imposed a requirement of at least three flights in each direction for each aircraft equipment type. In the later calculation of minimum, excess, and total flying time, if a flight had no available wind adjustment, it was discarded from the calculation. This minimum sample size requirement forces airports with few flights between them to be ignored in the computations, biasing our overall result towards larger, busier airports. In order to reduce this bias, we grouped airports by closeness to each other.[§] Wind adjustment values were computed between groups of airports ("clusters"), instead of between pairs of individual airports. Figures 3 and 4 show two different clusterings of the 3628 airports seen in our 30-day population.

Varying the number of clusters generates a trade-off decision between number of flights represented and fineness of the wind adjustment applied: more clusters means fewer airports per cluster, leading to finer wind adjustment, but also means discarding flights between smaller airports.

The resultant flying time metrics had minimal change as a function of number of clusters. The trade-off decision then hinged on number of flights represented. Analysis showed that 25 clusters captured a large number of flights. Therefore, wind adjustments were computed based on the 25 airport clusters.

As a check that wind effects were actually being captured with the clustering approach, Figure 5 shows a plot of angle of flight through the NAS versus speed adjustment, using 25 clusters. As expected, a rough sine wave pattern is seen, with high positive speed adjustment values ("speed-up") for west-flying flights (270 degrees), negative speed adjustment values ("slow-down") for east-flying flights (90 degrees), and small and zero adjustments for north- and south-flying flights (0,180, and 360 degrees). The range of speed adjustments for a given 10-degree bin are caused mostly by different equipment types. Figure 5 uses "box and whisker" elements for subsets of 10-degree bins. (See Tukey[8] for a full description of this display technique.)

[§] Using a popular clustering algorithm known as K-means.[9]

Using the comparison of dual distributions as before in Figures 1 and 2, it was determined that clustering of airports yielded good accuracy in adjusting speeds.

<u>Mechanics of Computing Flying Time Metrics</u>

Using ETMS flight position reports, which are based on automatic flight tracker output, flying times were computed by interpolation of the data, which are reported every minute or every 4 minutes (earlier in 2001). Next, as explained, a wind effect adjustment (specific to origin/destination cluster and aircraft equipment type) was made for each flight. For the remainder of this paper, when we refer to "flying times" we mean "adjusted flying times," i.e., flying times that have undergone the wind adjustment and have satisfied our minimum sample size requirements.

A "minimum flying time" was calculated for each origin/destination airport and aircraft equipment type combination. We required at least three flights per combination (otherwise a minimum value is not very meaningful) otherwise the flights were discarded from the calculation. The base value selected as the minimum flying time for comparison purposes was the *shortest* flying time for the origin/destination/aircraft equipment type combination. An analysis was performed to see if an alternate base value, e.g., fifth percentile, might be more appropriate. Examination of excess flying times showed the cumulative distribution function rose steeply and smoothly. Thousands of flights had excess flying times less than 30 seconds above the minimum. Based on this, it was decided that the shortest, the observed minimum value, was an appropriate base value for computing excess flying time.

Note that whereas the wind-adjustment factors were computed using *cluster* pairs, the flying time computation was based on *airport* pairs.

Next, using the collection of flight times for an origin/destination/aircraft equipment type combination and the observed minimum, an excess flying time for each flight could be computed via subtraction. By this means, the minimum time flight (or flights—it need not be unique) was considered to have an excess flying time of zero. The excess attributed to other flights was their total flying time minus the minimum of the corresponding origin/destination/aircraft equipment type combination. Total flying time, for the traversal from 100 nmi from origin to 100 nmi until destination, was also computed.

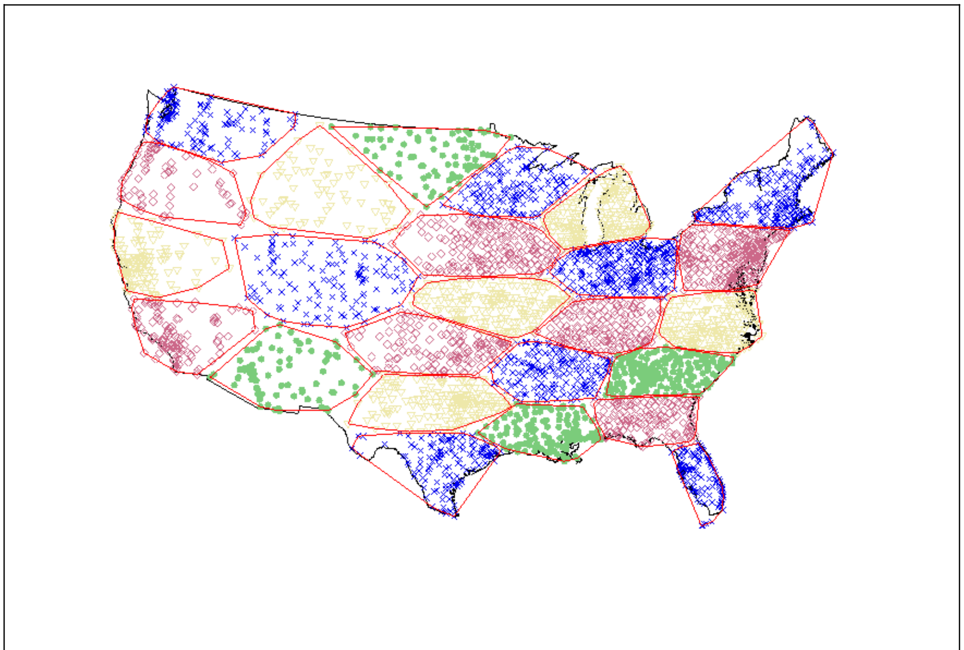

Figure 3. Airports Grouped into 25 Clusters

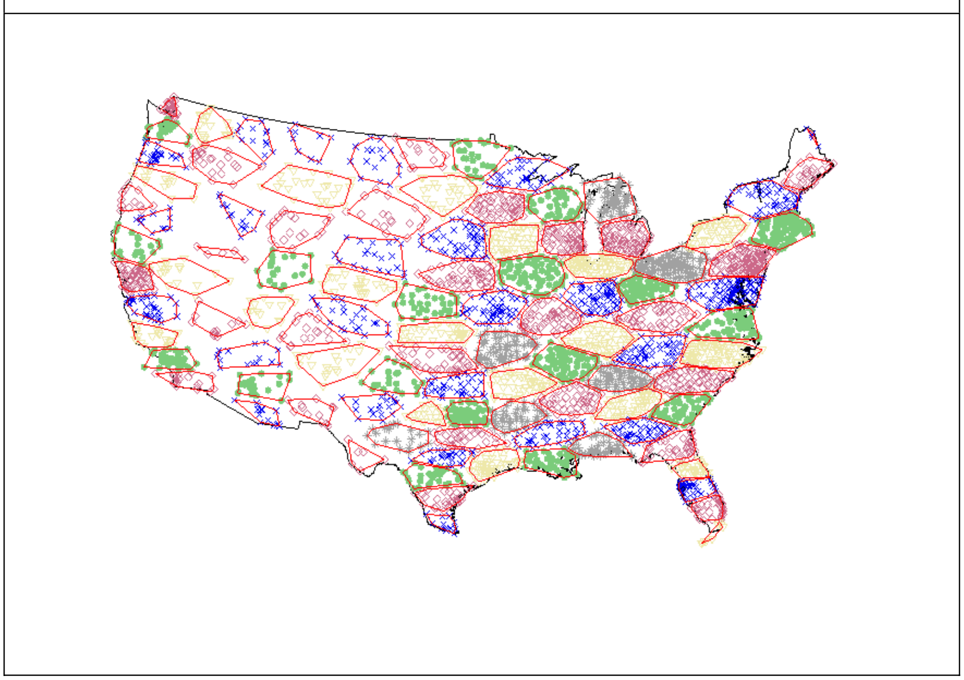

Figure 4. Airports Grouped into 100 Clusters

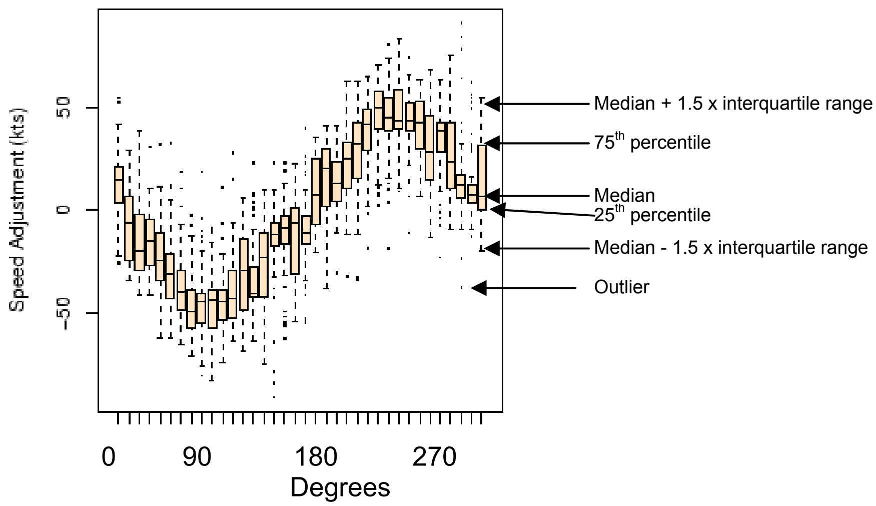

Figure 5. Computed Speed Adjustment Versus Angle of Flight Through the NAS Wind Field

We analyzed two 15-day sets of data, one for 2001 and one for 2002. For each of the two sets of days, a single minimum flying time was computed for each origin/destination/aircraft equipment type combination (i.e., one minimum for 2001 and another for 2002). The effect of analyzing a set of days is as follows. On any given day, many factors work together to influence how well the system performs overall. The minimum flying time possible on one day may be different from that on another because of a variety of factors, e.g., varying levels and locations of congestion, differing ATM strategies, etc. Therefore, as the number of days in a set of days is increased, one expects that, overall, the minimum flight times will decrease (since the opportunity grows to find a new, low minimum time as the sample size increases)—sharply at first and then more modestly, eventually approaching lower bounds (one for each origin/destination/aircraft equipment type combination). Hence, excess flying times will increase with the number of days in a set. The average excess flying time will approach an upper bound. For our 15-day samples per year, an analysis showed that the average excess flying times were at or near this upper bound.

Comparing 2001 to 2002

Operational data from the NAS has many sources of variability. To control variability, we grouped observations by origin airport, destination airport, and aircraft equipment type combinations. We compared 2001 metrics to 2002 metrics by aggregating within-combination differences. The Total Flying Time and Excess Flying Time metrics were tested using within-combination differences of means. As described in Appendix A, this was done using two-stage, customized statistical tests that depend on properties of means and variances.

Since tests to compare 2002 with 2001 for the minimum flying time and standard deviation of total flying time metrics do not directly involve differences of means, the customized tests mentioned above are not applicable. For these metrics, we used a simpler approach of weighting each origin/destination/aircraft equipment type combination by 1, subtracting 2001 from 2002 values, and testing if the mean of the combinations distribution was equal to zero. Additional analysis could be done to weight the minimum flight time differences and standard deviation of total flying time per combination to reflect the numbers of flights in each combination in 2002 and in 2001.

Results

At the NAS-wide level, results are in Table 1. As an interpretation of these metrics, one might say that the best flights are arriving sooner (minimum flights are shorter) in 2002, and average total flying time is 23 seconds shorter in 2002. The standard deviation of the total flying time is 6 seconds shorter in 2002.

How important is this rather small difference? When considered in light of the number of flights per day in the NAS (40–50 thousand flights per day receive ATC services), a total flying time difference of 23 seconds becomes significant. We did not attempt to determine the causes of the improvement. One might guess it is a combination of lower traffic congestion in 2002 and the influence of automation and procedural programs of the Operational Evolution Plan (OEP).[10]

The distribution of the differences of minimum flying times, combination by combination, is shown in Figure 6. It is symmetric, as expected, but has more of a central "peak" than a normal distribution. The spread or dispersion shows there was a range of differences from about -10 to +10 minutes. The other metrics are similar in their distributions of by-combination differences.

Also examined were 22 city pairs of interest. The results are in Table 2. The total flying time metric is significant, with a value of 32 seconds shorter in 2002. The average excess flying time was also lower, by 26 seconds, in 2002. The per-combination minimum flying times were likewise lower in 2002.

More detailed results are described next for 22 city pairs. Figure 7 shows the total flying time difference for 22 city pairs (per X-axis label). A total of 44 "box-and-whisker" elements are shown (with whiskers suppressed), since flights were analyzed separately for each direction. For example, leftmost in Figure 7, the box pair labeled ATL-DFW (Atlanta Hartfield International Airport/Dallas-Ft. Worth International Airport) are flights from ATL to DFW (the left one of the pair) and flights from DFW to ATL (the right one of the pair). Note the differences in the city pair-specific distributions. For some city pairs, there is an increase in 2002, and for others a decrease. Most of the observations in the figure are below the zero line, corresponding to lower flying times in 2002.

Further detailed analysis is possible to pinpoint the causes of city-pair-specific differences. It may be that certain congestion points in en route airspace are causing inordinate increases in flying time. Identification of these points could lead to the amelioration of the problem. For example, re-routing and improved automation and procedures could be used to reduce flying times through current congestion points.

Drilling down further, Figures 8 and 9 show the excess flying time for flights from Atlanta to Dallas/Ft.Worth flying the MD-80 aircraft type. Each distribution is bounded below by zero and has a right tail. The average value for 2001 in Figure 8 is 4.37 minutes, and for 2002 in Figure 9 is 4.06 minutes, a difference of 0.31 minutes. Sample sizes are 98 and 142 flights, respectively.

Considering Sources of Variation

As mentioned earlier, there are multiple sources of variation in our problem structure, including winds, aircraft equipment type, and origin/destination airport pair. We performed experiments to see how our results would have changed if we had not performed wind corrections and not considered aircraft equipment type. The metric tested was the total flying time, using the simpler unweighted comparison of averages.

Table 3 shows the results of these experiments, essentially a two-squared factorial setup with aircraft equipment type off/on and wind adjustment off/on. Accounting for both sources of variation is obviously the correct approach, but Table 3 shows this numerically: it delivers the most significant result. These results also show that, of the two, wind adjustment is a larger source of variation than aircraft equipment type—if winds are ignored then the test significance (testing the null hypothesis that the mean is zero) is large, either 0.71 or 0.89. One might possibly ignore aircraft equipment type to obtain a rough-cut answer, but ignoring the winds guarantees a wrong conclusion of "no difference."

Error Analysis

Operational data are subject to various sources of error. For this study, we have identified three potentially signficant sources of error and need to assure ourselves that our results are valid in light of the potential error. The three sources of error are: (1) aircraft position reports at latitude/longitude contain degrees and minutes, but not seconds, (2) timestamp data for aircraft position reports were truncated to 1-minute precision prior to our processing, and (3) linear interpolation was used to find the time of crossing the 100 nmi (from origin and destination airport) demarcation lines.

Table 1: NAS-Wide Flying Time Metrics, Per Flight, 2002 Versus 2001

Metric	Difference	Signficance
Total Flying Times	Average was 23 seconds shorter in 2002, Standard deviation was 6 seconds smaller in 2002.	p<0.0001 p=0.0001
Minimum Flying Times	Average was 26 seconds shorter in 2002.	p< 0.0001
Excess Flying Times	Not significant.	p=0.24

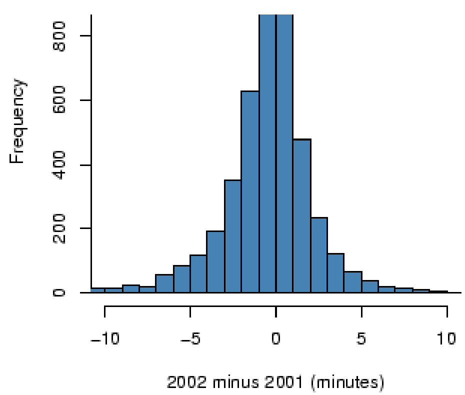

Figure 6: Minimum Adjusted Flying Time Paired Differences per Combination Empirical Distribution, 2002 Minus 2001

Table 2: Twenty-two City Pairs Flying Time Metrics, 2002 Versus 2001

Metric	Difference	Signficance
Total Flying Times	Average was 32 seconds shorter in 2002. Standard deviation was 18 seconds smaller in 2002.	p<0.0001 p=0.004
Minimum Flying Times	Not signficant.	p=0.08
Excess Flying Times	Average was 26 seconds lower in 2002.	p<0.0001

Table 3: Comparing Results, Ignoring Different Sources of Variance

Aircraft Equipment Type Considered?	Wind Adjustment Performed? No	Wind Adjustment Performed? Yes
No	Difference = 4 seconds p = 0.89	Difference = 28 seconds p = 0.02
Yes	Difference = 5 seconds p = 0.71	Difference = 22 seconds p < 0.0001

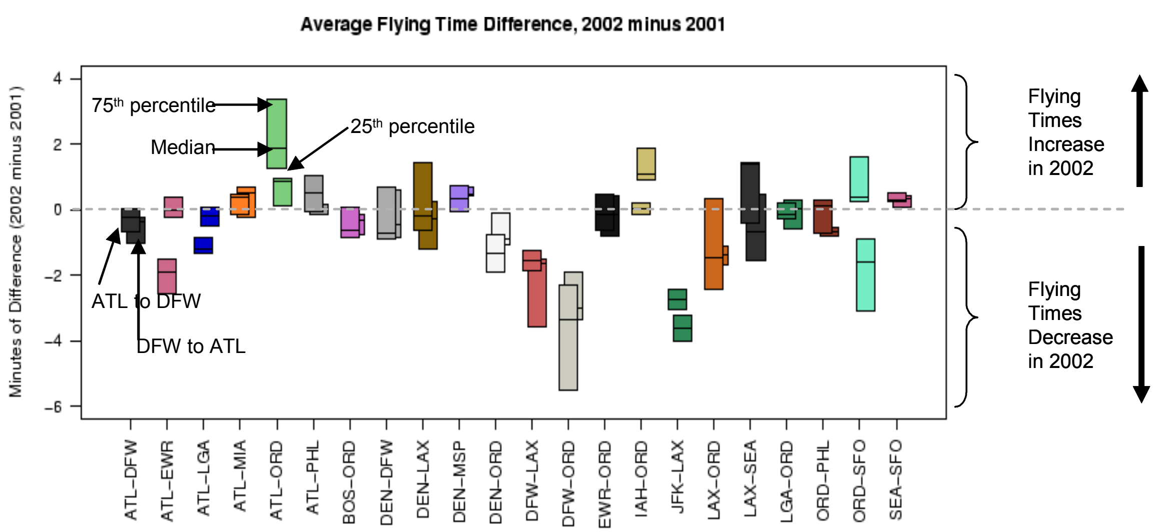

Figure 7: Total Adjusted Flying Time Difference, 2002–2001, for 22 City Pairs

To assess the impact of error, experiments were performed that perturbed the position or time data, and examined sensitivity of response in the total flying time difference value. This difference was not the weighted means test employed earlier, rather it was a simpler scheme of weighting each origin, destination, and aircraft equipment type by one, and taking the difference, by combination, 2002 minus 2001. Specifically, we looked at several treatments one at a time, and then combined treatments. Treatments are described in Table 4.

Computer runs were executed using the various treatments, and the final result, the difference 2002-2001 total flying time, was assessed. Table 5 shows these results. All results are within the 95 percent confidence bound on the "baseline" runs and are deemed acceptable. Other experiment runs which combined treatments obtained similar results. We conclude that the results of this paper hold up even in light of these errors in the data.

Conclusions and Next Steps

We have used operational aircraft track data to compute flying time metrics so that the years 2001 and 2002 could be compared. The aircraft tracks were also used to remove, to a large degree, the influence of winds, using the assumption that flights in opposite directions experience opposite winds. Our analysis found that, for our 30 good weather days, flying times were lower in 2002 compared to 2001.

The paired-observations approach, which matched categories of origin airport, destination airport, and aircraft equipment type, was useful for removing variability and creating a fair comparison between the years.

Furthermore, we examined 22 selected city pairs and found a wide variation in the differences in flying times across the years. Determining the reasons for these specific differences would be interesting follow-on work. Other follow-on work should consider bad weather days in an attempt to find a fully-annualized, comparable flying time metric.

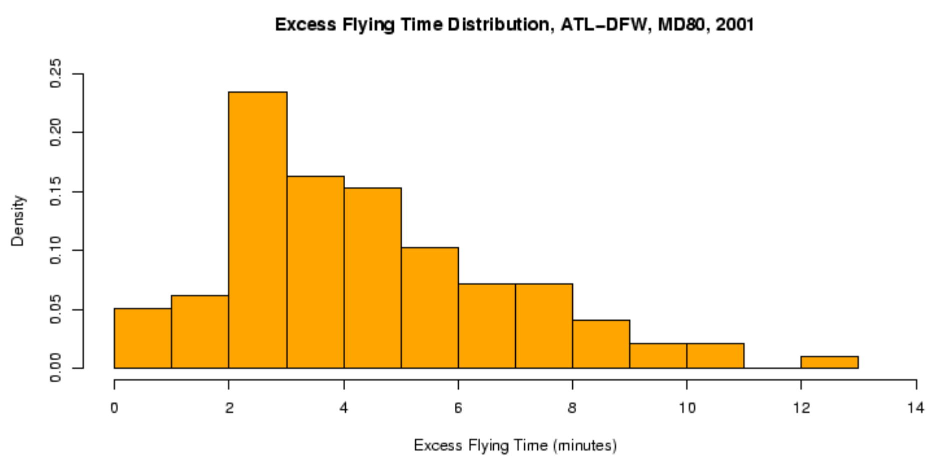

Figure 8: Excess Flying Time Distribution, ATL-DFW, MD80, 2001

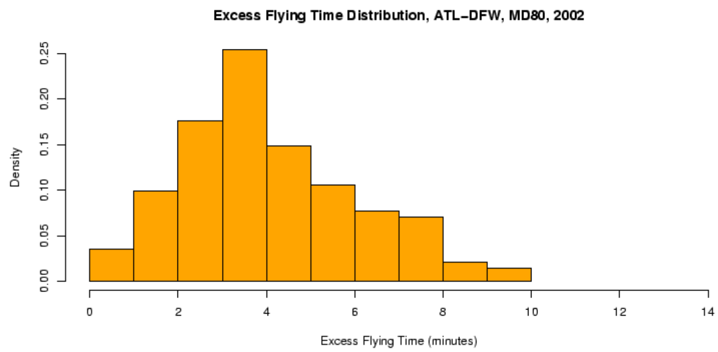

Figure 9: Excess Flying Time Distribution, ATL-DFW, MD80, 2002

Table 4: Experimental Treatments for Assessing Suspected Error Sources

Treatment	**Error Source**	**Experiment**
1	Aircraft position latitude and longitude were assumed *rounded* to degrees and minutes, losing the precision of seconds.	Add a random value between -0.5 and 0.4999 minutes to latitude, and separately, to longitude, to mimic "un-doing" the effect of rounding. Three runs were performed, with different random number seeds.
2	Aircraft position latitude and longitude were assumed *truncated* to degrees and minutes, losing the precision of seconds.	Add a random value between 0 and 0.9999 minutes to latitude, and separately, to longitude, to mimic "un-doing" the effect of truncation. Three runs were performed, with different random number seeds.
3	Aircraft position timestamps were truncated to minutes prior to use, losing the precision of seconds.	Add a random value between 0 and 0.9999 minutes to the time value to mimic "un-doing" the effect of truncation. Three runs were performed with different random number seeds.
4	Alternate interpolation algorithm #1. The original interpolation algorithm used two points which spanned the 100 nmi demarcation ring. This variant used an additional position report for smoothing. Linear interpolation is used.	No randomization necessary, simply apply alternate algorithm.
5	Alternate interpolation algorithm #2, using 3 position reports and assuming constant acceleration. This method computes the time that the aircraft is at a given distance from an origin when the time and distance of the aircraft are known at three points. This is done using the equation of motion for each of the three known points to compute the constant acceleration and the distance and speed at time zero. The computed values are substituted into the equation of motion for the given distance. This equation is then solved to determine the time at which the aircraft was at the given distance from the origin.	No randomization necessary, simply apply alternate algorithm.

Table 5: Results of Treatments Assessing Suspected Error Sources

Baseline	Treatment 1			Treatment 2		
	Seed1	**Seed2**	**Seed3**	**Seed1**	**Seed2**	**Seed3**
-20.29 * **-23.32** -26.34 **	**-23.33**	**-23.77**	**-24.02**	**-23.38**	**-23.82**	**-22.71**

	Treatment3			Treatment4	Treatment5
Baseline	**Seed1**	**Seed2**	**Seed3**		
-20.29* -23.32 -26.34**	**-22.40**	**-23.20**	**-21.70**	**-23.33**	**-23.38**

*Lower 95% confidence bound.

** Upper 95% confidence bound.

References

1. Bolczak, C., L. Brown, J. Hoffman, C. Schellenberg, and W. Voss, 1997, "Flexibility, Predictability, and Access: An Expanded View of System Performance," *ATC Quarterly*, Arlington, VA: ATCA Institute.
2. Ball, M., J. DeArmon, J. Pyburn, 1995, "Is Free Flight Feasible? Results from Initial Simulations," *Journal of Air Traffic Control*, Arlington, VA: ATCA.
3. Willemain, T., 2001, "Estimating Components of Variation in Flight Times," http://www.its/berkeley.edu/nextor/pub/WP-01-2.pdf.
4. Bureau of Transportation Statistics, 2002, Airborne Flight Time Analysis, http://www.bts.gov.ntda/oai/DetailedStatistics/
5. Bureau of Transportation Statistics, 2002, Office of Airline Information, http://www.bts.gov/oai/.
6. Callaham, M. et al., 2001, "Assessing NAS Performance: Normalizing for the Effects of Weather," Fourth USA/Europe Air Traffic Management R&D Symposium, Santa Fe, NM.
7. Oiesen, R. 2000, "The ETMS Hub Site and Weather," http://www.caasd.org/library/presentations/weather_conference/2_Oiesen.pdf.
8. Tukey, J., 1977, *Exploratory Data Analysis*, Addison-Wesley, Reading, PA: McGraw Hill.
9. Hartigan, J., 1976, *Clustering Algorithms,* New York: Wiley and Sons, Inc.
10. Federal Aviation Administration, 2002, Operational Evolution Plan Description, http://www2.faa.gov/programs/oep.
11. Walpole, R., et al., 2002, *Probability and Statistics for Engineers and Scientists*, Upper Saddle River, NJ: Prentice Hall.

Appendix A: Customized Tests for Total Flying Time and Excess Flying Time Metrics

For the Total Flying Time and Excess Flying Time metrics, for each combination, indexed by i, we computed the difference, d_i, of the 2001 mean and the 2002 mean. We computed the sample variance, v_i, and degrees of freedom, n_i, for each difference using formulas for the T-test of a difference of means when the population variances are unknown and different. (See tests concerning means and variances in Walpole and Myers.[11]) Since degrees of freedom are the effective number of observations, we computed the *weighted average difference* of the combinations, d, using weights proportional to the degrees of freedom. That is, if $N = \Sigma n_j$, (j is a dummy index for the combinations), and $w_i = n_i /N$, then the w_i are the weights and $d = \Sigma(w_i * d_i)$.

Using standard properties of variances (see Walpole and Myers[11]) and assuming that the random variables corresponding to the d_i are mutually stochastically independent, we estimated the variance, V, of the weighted average difference as

$$V = \text{Variance}[\Sigma(w_i * v_i)] = \Sigma(w_i^2 * v_i)$$

By the Central Limit Theorem, under the null hypothesis that the random variable, D, corresponding to d equals zero, the statistic D / Sqrt(V) is distributed approximately as a standard normal random variable. Therefore, if the absolute value of d / $\text{Sqrt}[\Sigma(w_i^2 * v_i)]$ exceeds 1.96 (which is $z_{.025}$) the null hypothesis that the weighted difference of means is zero is rejected at level .05 in favor of the two-sided alternative hypothesis that the 2001 and 2002 values differ.

A p-value is presented in the results section.

NOTE: The contents of this material reflect the views of the authors and/or Director of the Center for Advanced Aviation System Development. Neither the Federal Aviation Administration nor the Department of Transportation makes any warranty or guarantee, or promise, expressed or implied, concerning the content or accuracy of the views expressed herein