ON-TIME PERFORMANCE

ASPM NON-PARAMETRIC STATISTICAL ANALYSIS

ACTUAL_OFF minus ACTUAL_OUT
ACTUAL_OFF minus SOBT

August 13, 2018

Tom Kozon

Introduction

ANALYSIS INTRODUCTION

NASA'S Airspace and Technology Demonstration 2 (ATD-2) project is currently aimed at integrating arrival, departure and surface operations to develop new capabilities for operational benefits. Such benefits include the improvement of flight efficiency and reduced emissions while maintaining or improving other factors such as operational throughput. The purpose of this analysis is to provide a data-driven examination of two selected flight metrics, i.e., (1) total taxi-out time and (2) actual off time – scheduled off block time. More specifically, ASPM data were analyzed to determine any possible differences in CLT departure flights on these two metrics when comparing pre-IADS against post-IADS metering operations.

Metrics and Data

METRICS EXAMINED AND DATA USED FOR ANALYSIS

Two metrics were selected for analysis, namely:

1) Actual-Off - Actual-Out as a measure of total taxi-out time. To reduce the chance of spurious data confounding the analysis, only those CLT departure flights with total taxi-out times greater than 2 minutes and less than or equal to 120 minutes were included in the analysis

2) Actual-Off - Scheduled Off Block Time (SOBT). Only those CLT departure flights with calculated values greater than 2 minutes and less than or equal to 120 minutes were included in this analysis.

Analysis of each of the two above-named metrics used the following 3 samples of CLT departure flights, for three analyses per metric, or six total analyses:

1) airport-wide all flights
2) airport-wide bank2 flights only
3) airport-wide all flights except bank2 and bank3 flights.

Aviation System Performance Metrics (ASPM) data were used for the analysis of the metrics of interest. CLT departure flights from 2016 through 2018 were extracted from the ASPM database to capture the trend from 2016-2018 on a yearly basis. The intent is to ascertain differences between the years as a possible method to determine if the introduction and implementation of IADS metering might have any impact on the metrics of interest. To reduce the chance of confounding factors affecting the analysis, only January through April data were used, since the trigger-point that determines the actual-out time changed in May 2018 to a laser-based system (DGS). It would also make analytic sense to use the same months for analysis across the three years.

Analytic Method

ANALYTIC METHOD USED

Initial examination of the data indicated that the distributions for each of the three years within each of the six analyses were negatively skewed. Since we could not assume normality in each of these distributions, a statistical analysis procedure with non-parametric assumptions was implemented.

We selected the Kruskal-Wallis procedure to test the statistical differences between the three years for each analysis. Hence, it was possible to calculate, for each of the six analyses, one omnibus ANOVA result and three post-hoc pairwise comparisons.

For each analysis, we tested the null hypothesis as follows:

H_0 There are no statistical differences in the distributions between the three years.

If the null hypothesis was rejected, then the 3 pairwise comparisons were examined for statistical significance.

Results Summary (supporting data presented later)

RESULTS SUMMARY
(supporting data presented later in this document)

Six Omnibus Anova Results

The null hypothesis was rejected consistently for all six analyses:

1) airport-wide all flights
   a) Actual_off– Actual_out
   b) Actual_off – SOBT

2) airport-wide bank2 flights
   a) Actual_off– Actual_out
   b) Actual_off – SOBT

3) airport-wide all flights except bank2 and bank3 flights
   a) Actual_off– Actual_out
   b) Actual_off – SOBT

Eighteen Pairwise Comparisons

The null hypothesis was rejected with most pairwise comparisons.

Results and Discussion (supporting data presented later)

What the data show

The distributions mostly show a very small upward trend across the three years in the mean values, usually no more than about 1 minute overall. While the computed values on the DVs were a bit higher in 2018, there was also **a larger proportion of regional flights to all other flights in 2018** which probably contributed to the slightly larger mean taxi-out time. This might also reflect special issues faced with concourse E which is occupied mostly by regional flights.

It should also be mentioned that **the median values tend to be very close to each other across the three years, and with one analysis, they are exactly the same.**

Why the null hypothesis was largely rejected

Because of the large number of statistical observations, the degrees of freedom in the computational procedure are elevated, so even very small differences in the 3 distributions will result in a statistically significant difference. This will have a large impact on how these results should be interpreted.

What this probably means

Differences between the 3 flight distributions (2016 2017 2018) on each dependent measure might be statistically significant but not meaningful(e.g., total mean taxi-out time differences of ~ 1 minute). Also, these differences will likely reflect factors other than IADS metering, such as the relatively large number of regional departure flights during 2018 which elevated the taxi-out time (and other possible covariates, e.g., weather).

Hence, it seems likely that ***the "DO NO HARM" standard was met***. Also, analysis results may suggest possible ***taxi-out improvement in 2018 (described later***).

Kruskal-Wallis Non-Parametric ANOVA

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of OFFminusOUTmin_AirportWide AllFlights is the same across categories of DEP_YEAR.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.
2	The distribution of OFFminusOUTmin_AirportWide Bank2 is the same across categories of DEP_YEAR.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.
3	The distribution of OFFminusOUTmin_AirportWide_NO_Bank23 is the same across categories of DEP_YEAR.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.

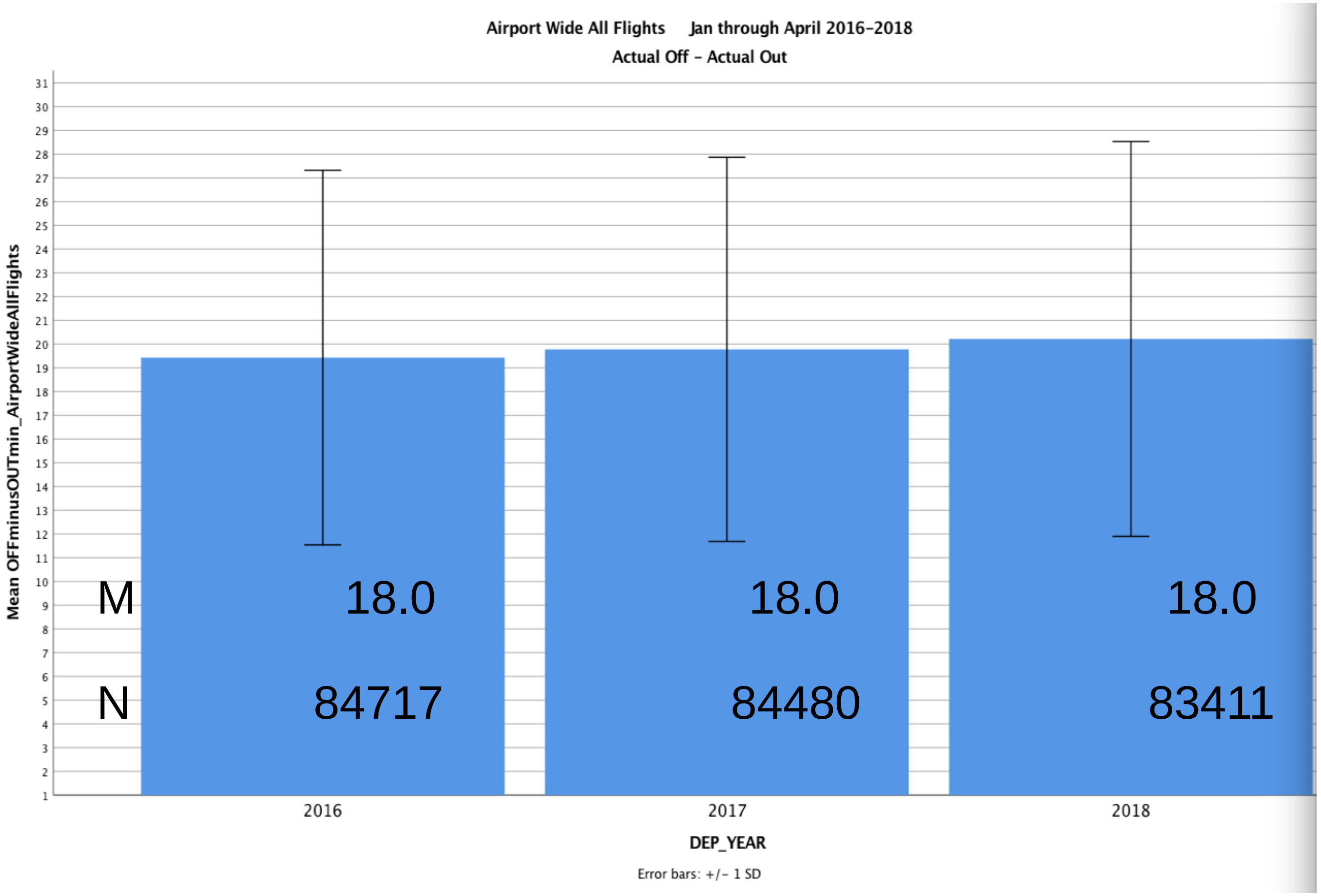
Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of OFFminusSOBTmin_AirportWide AllFlights is the same across categories of DEP_YEAR.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.
2	The distribution of OFFminusSOBTmin_AirportWide Bank2 is the same across categories of DEP_YEAR.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.
3	The distribution of OFFminusSOBTmin_AirportWide_NO_Bank23 is the same across categories of DEP_YEAR.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.

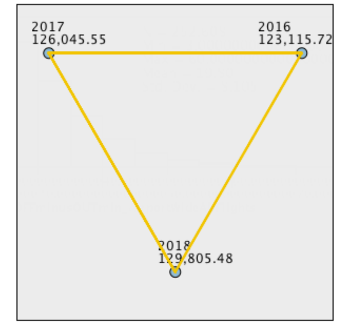
Asymptotic significances are displayed. The significance level is .05.

All omnibus anovas for each of the six analyses yielded results that reject the null hypothesis.

Off-Out Airport Wide All Flights



Pairwise Comparisons of DEP_YEAR



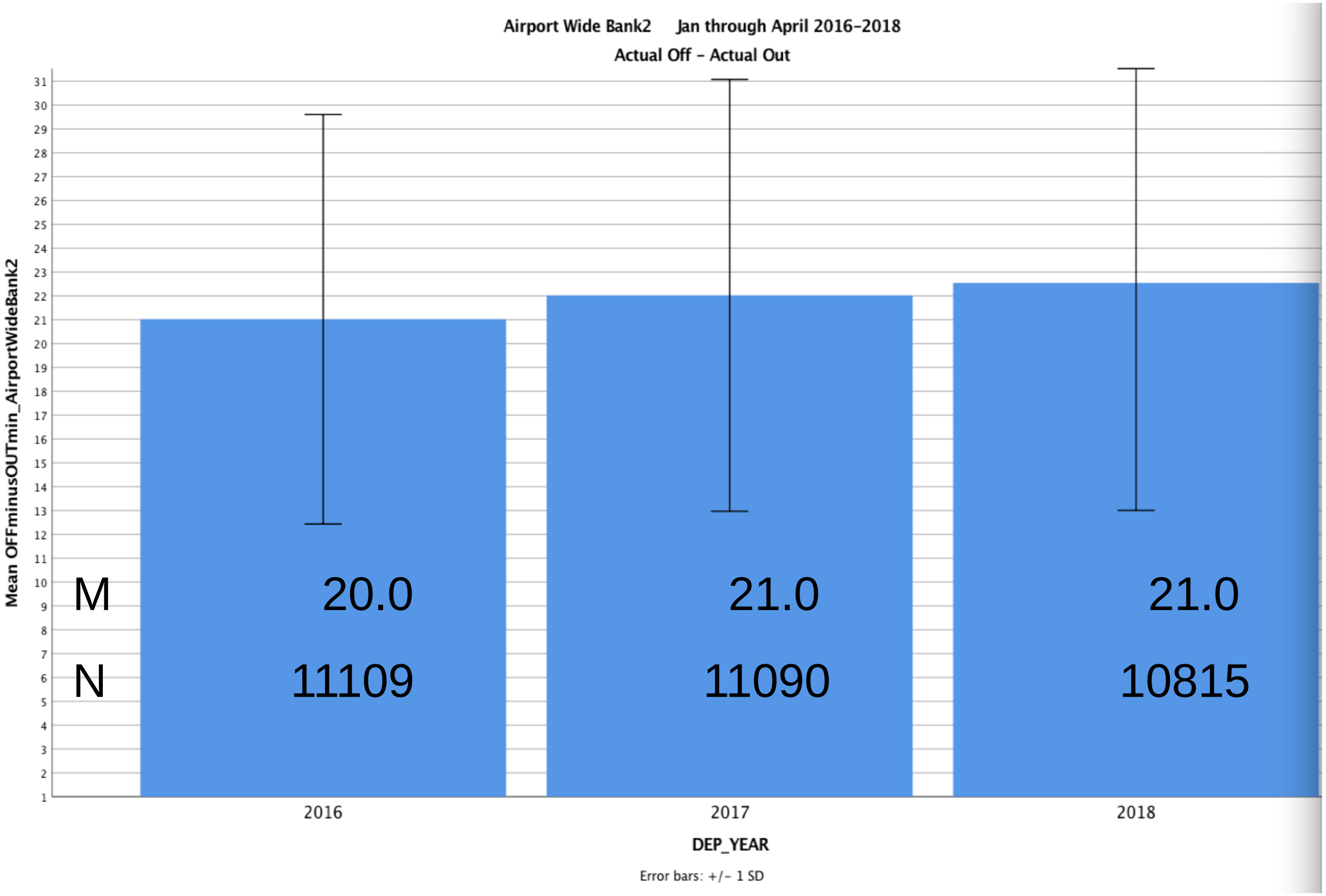
Each node shows the sample average rank of DEP_YEAR.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
2016-2017	-2,929.829	354.194	-8.272	.000	.000
2016-2018	-6,689.762	355.329	-18.827	.000	.000
2017-2018	-3,759.934	355.576	-10.574	.000	.000

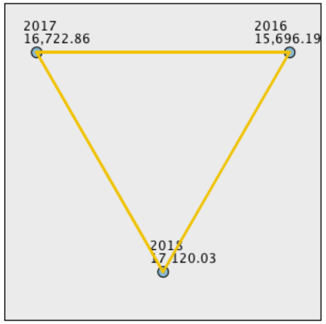
Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.
Significance values have been adjusted by the Bonferroni correction for

While the null hypothesis was rejected very small differences between the three years in the mean values were observed and the median values are exactly the same. Hence, this statistically significant difference is not a meaningful difference, i.e., in practical terms, the three distributions are the same.

Off-Out Airport Wide Bank2



Pairwise Comparisons of DEP_YEAR



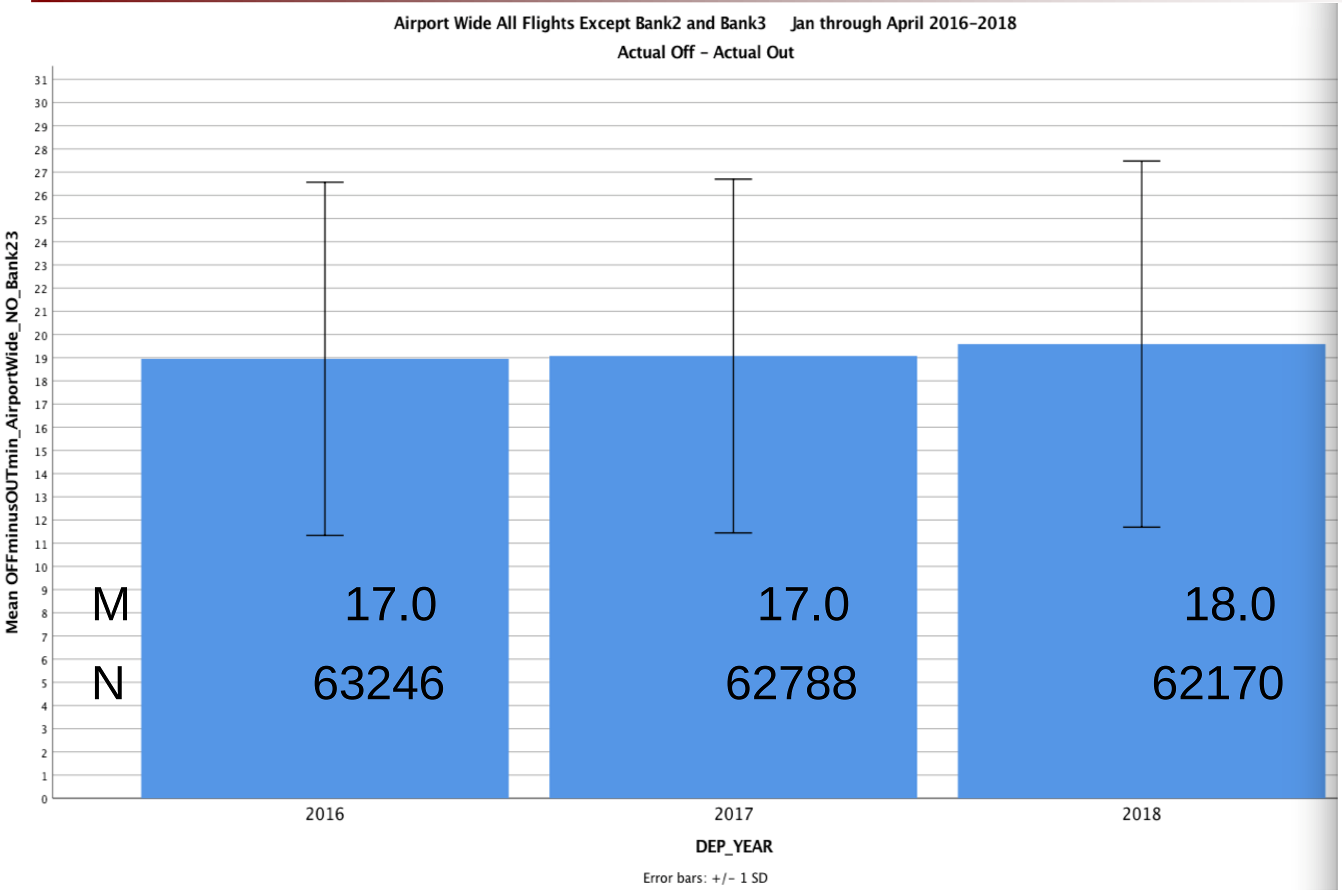
Each node shows the sample average rank of DEP_YEAR.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
2016-2017	-1,026.674	127.842	-8.031	.000	.00
2016-2018	-1,423.849	128.653	-11.067	.000	.00
2017-2018	-397.174	128.707	-3.086	.002	.00

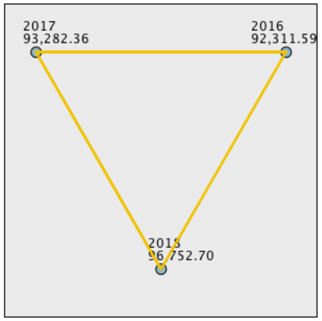
Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.
Significance values have been adjusted by the Bonferroni correction for multiple tests.

As with the Airport-Wide All-Flights analysis, the null hypothesis was rejected, very small mean differences between the three years were observed and the median values are nearly the same. Hence, this statistically significant difference is also not a meaningful difference

Off-Out Airport Wide All Flights Except Bank2 and Bank3



Pairwise Comparisons of DEP_YEAR



Each node shows the sample average rank of DEP_YEAR.

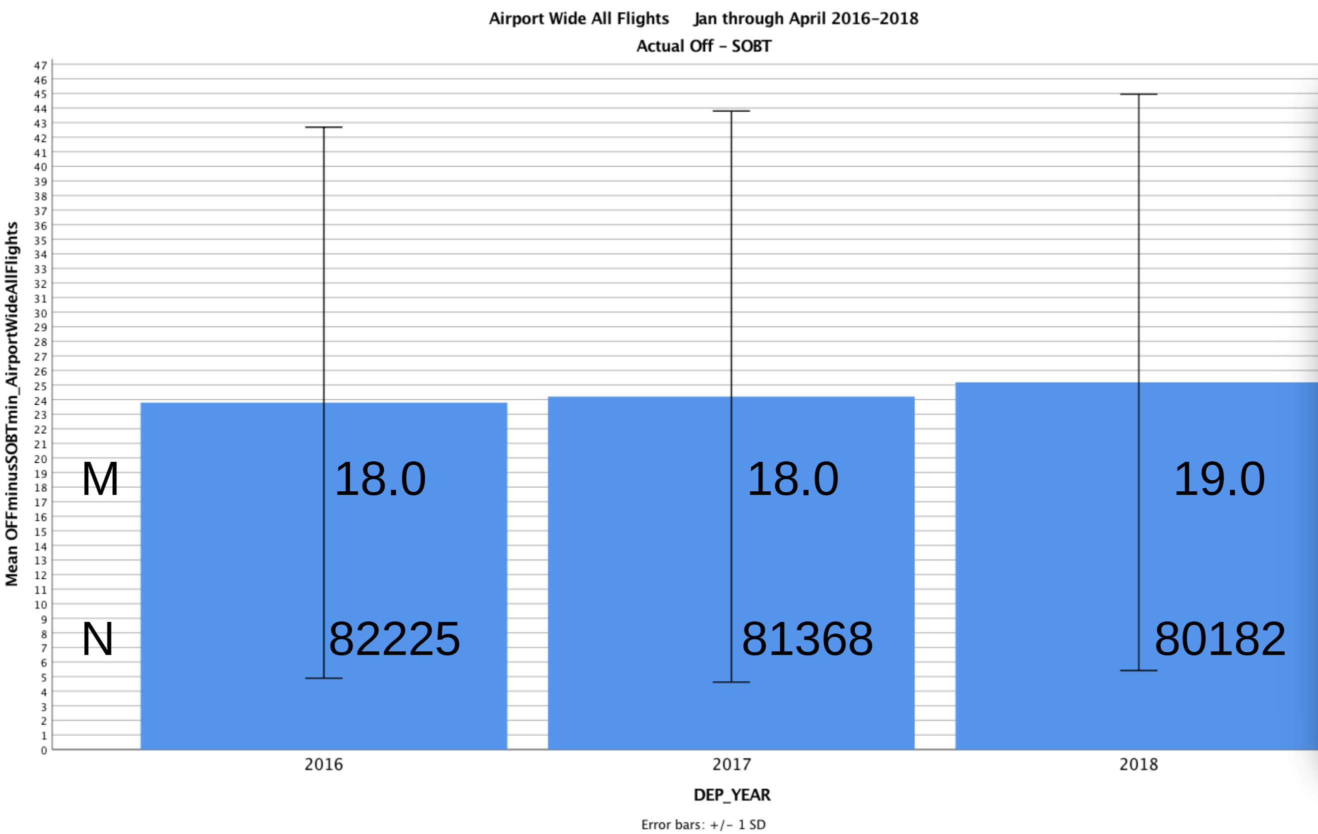
Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
2016-2017	-970.765	305.714	-3.175	.001	.004
2016-2018	-4,441.101	306.475	-14.491	.000	.000
2017-2018	-3,470.336	307.029	-11.303	.000	.000

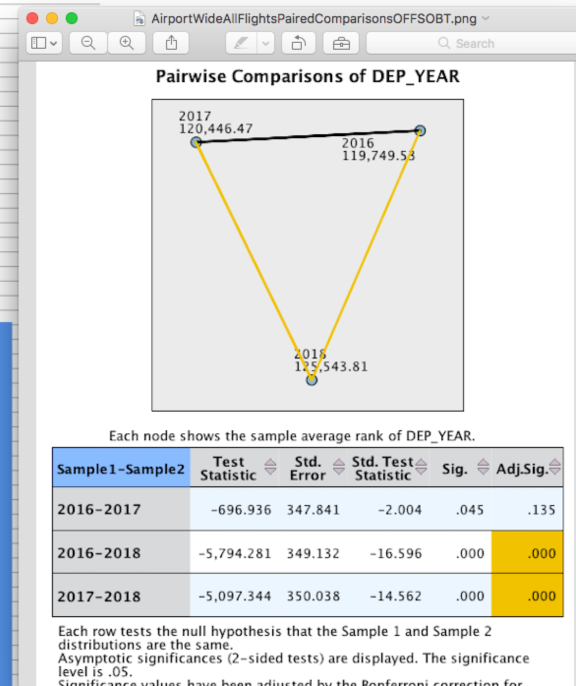
Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.
Significance values have been adjusted by the Bonferroni correction for multiple tests.

Again, the the null hypothesis was rejected, and in this analysis, extremely small mean differences between the three years were observed (a fraction of a minute) and the median values are nearly the same. Hence, this statistically significant difference is also not a meaningful difference.

Since this analysis only includes flights not subject to metering, it seems clear that the higher mean and median values in 2018 are not caused by IADS metering. Therefore, other factors that make 2018 flight operations different from 2016/2017 would help to explain this difference, for example, more regional flights in 2018.

Off-SOBT Airport Wide All Flights





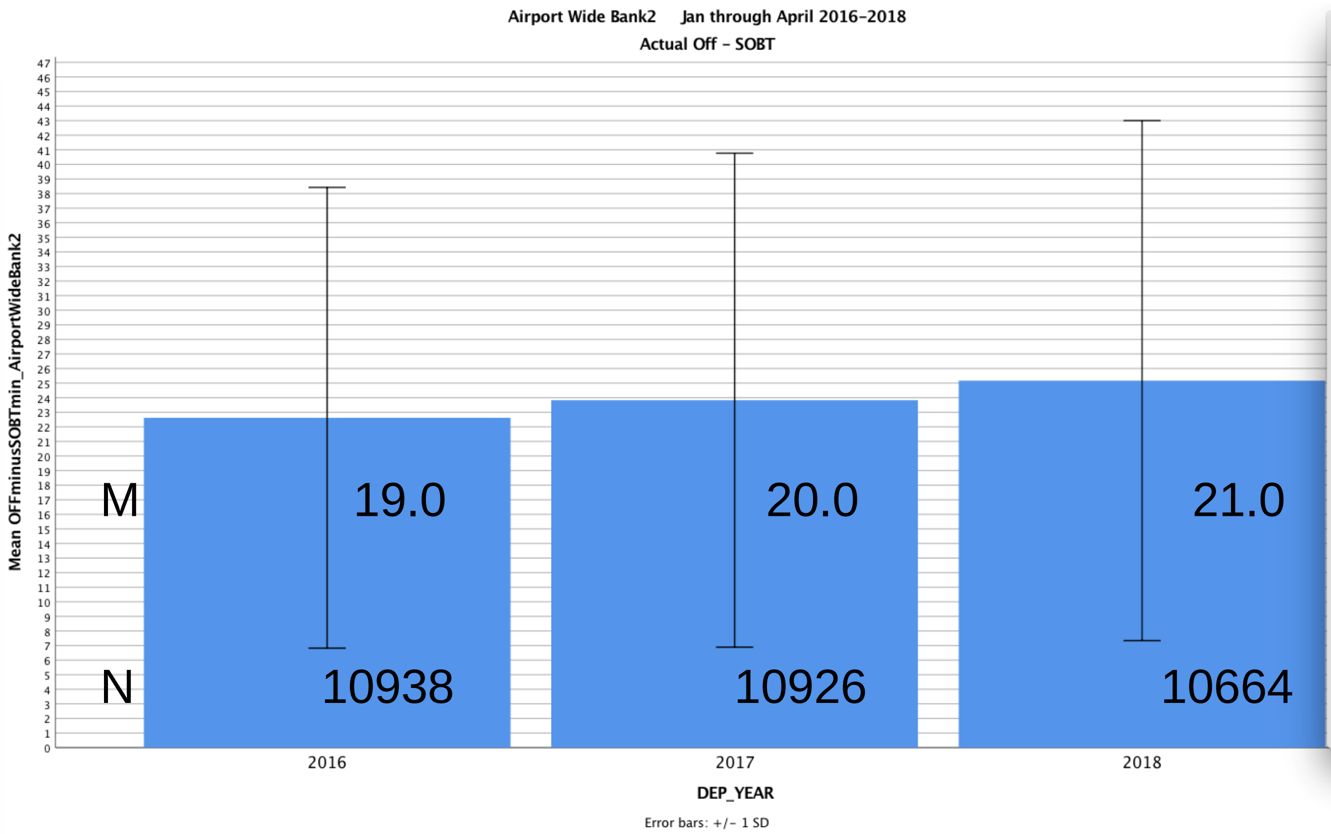
Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
2016-2017	-696.936	347.841	-2.004	.045	.135
2016-2018	-5,794.281	349.132	-16.596	.000	.000
2017-2018	-5,097.344	350.038	-14.562	.000	.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.
Significance values have been adjusted by the Bonferroni correction for

Statistically significant difference is not a meaningful difference.

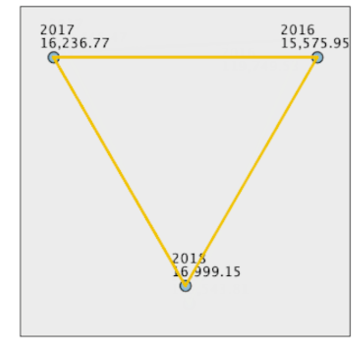
Off-SOBT Airport Wide Bank2

ATD2 Integrated Arrival/Departure/Surface



AirportWideBank2PairedComparisonsOFFSOBT.png

Pairwise Comparisons of DEP_YEAR



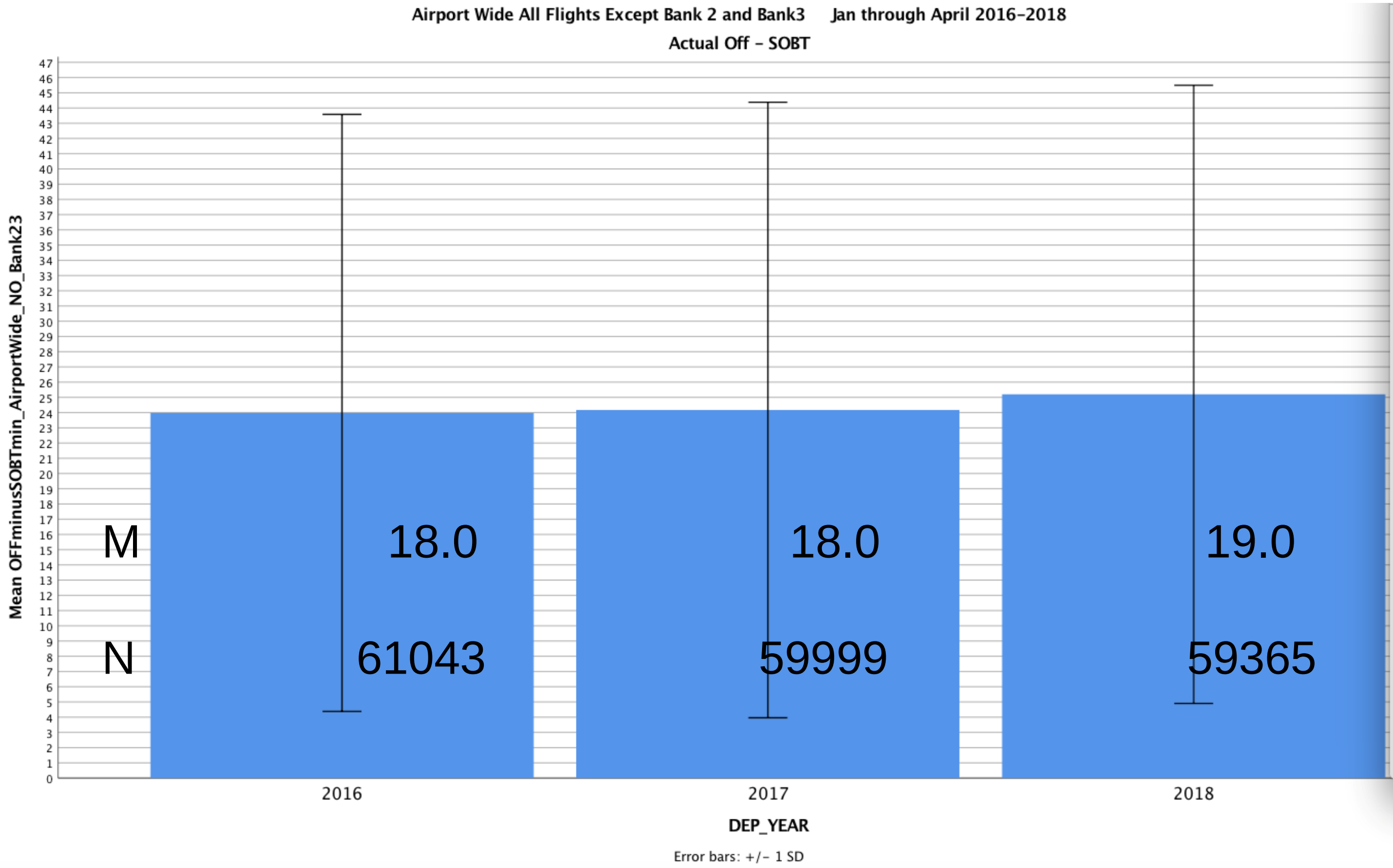
Each node shows the sample average rank of DEP_YEAR.

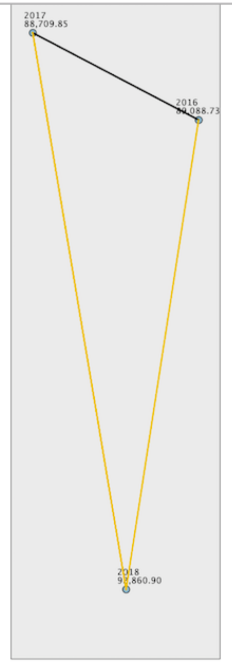
Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
2016-2017	-660.819	126.964	-5.205	.000	.000
2016-2018	-1,423.192	127.742	-11.141	.000	.000
2017-2018	-762.373	127.777	-5.966	.000	.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.
Significance values have been adjusted by the Bonferroni correction for multiple tests.

Statistically significant difference is not a meaningful difference.

Off-SOBT Airport Wide All Flights Except Bank2 and Bank3





Each node shows the sample average rank of DEP_YEAR.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
2017-2016	378.885	299.269	1.266	.206	.617
2017-2018	-4,151.057	301.359	-13.774	.000	.000
2016-2018	-3,772.172	300.074	-12.571	.000	.000

Statistically significant difference is not a meaningful difference. Also, since this analysis only includes flights not subject to metering, it seems clear that the higher mean and median values in 2018 are not caused by IADS metering. Therefore, other factors that make 2018 flight operations different from 2016/2017 would help to explain this difference, e.g., more regional flights in 2018.

Airport Wide: Relative Frequency of Regional to All Other Flights

DEP YEAR	OFF-OUT Airport Wide All	OFF-OUT Airport Wide Regional	Percentage of Regional to All
2016	**84717**	**37494**	**44%**
2017	**84480**	**37287**	**44%**
2018	**83411**	**39840**	**48%**
DEP YEAR	**OFF-SOBT Airport Wide All**	**OFF-SOBT Airport Wide Regional**	**Percentage of Regional to All**
2016	**82225**	**36554**	**44%**
2017	**81368**	**36130**	**44%**
2018	**80182**	**38354**	**48%**

Across the six analysis metrics, mean values in 2018 tend to be slightly larger than 2016/2017. One likely reason is the higher percentage of regional flights to all flights in 2018. Since regional flights usually occupy concourse E and also tend to have higher taxi-out times than other flights, it would make sense that taxi-out times would be *at least* slightly higher in 2018.

Selected Summary Factors and Conclusions

All six analyses yielded statistically significant, but not meaningful differences. Statistical significance occurred due to the large frequency of observations having the effect of elevating the degrees of freedom in the computational procedure. So, very small differences in the tested distributions will likely cause the rejection of the null hypothesis, even though the mean and median differences are not meaningful.

The two analyses testing the yearly differences using all airport-wide departure flights except those in Bank2 and Bank3 show an increase in mean and median values in 2018 as compared to 2016/2017. Since none of the flights are subject to IADS metering in this sample, any increase in 2018 is due to factors other than IADS metering, for example an increased proportion of regional flights to all flights in 2018.

Hence, it seems clear that the “do not harm” standard was met.

The total taxi-out time is defined as the actual-off time minus the actual-out time. ASPM actual-out times prior to May 1, 2018 are based on the brake-release time. Since IADS metering holds occur after this brake-release event, the taxi-out time for metered flights will include metering-hold time, i.e., times when the aircraft is not moving. So, taxi-out time for 2018 will be elevated if we define it from the time the aircraft moves away from the gate and not the brake-release event. Therefore, the calculated taxi-out times in this sample are increased in 2018 for this reason, i.e., a data artifact, and not because of increased time from moving away from the gate to the actual-off time. Hence the actual taxi-out times in 2018, defined as actual-off time minus the time the aircraft moves away from the gate, are likely smaller than those calculated in this sample.

Hence, it seems likely that there is improvement in taxi-out time in 2018 as compared to 2016/2017 even though it was not directly detected in this data sample.